Lecture Notes in Physics

V. G. Gurzadyan D. Pfenniger (Eds.)

Ergodic Concepts in Stellar Dynamics

Proceedings of an International Workshop
Held at Geneva Observatory
University of Geneva, Switzerland, 1-3 March 1993

Springer-Verlag
Berlin Heidelberg New York
London Paris Tokyo
Hong Kong Barcelona
Budapest

Editors

V. G. Gurzadyan
Department of Theoretical Physics
Yerevan Physics Institute
375036 Yerevan, Armenia

D. Pfenniger
Geneva Observatory
University of Geneva
CH-1290 Sauverny, Switzerland

ISBN 3-540-57929-X Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-57929-X Springer-Verlag New York Berlin Heidelberg

CIP data applied for

SPIN: 10080337 55/3140-543210 - Printed on acid-free paper

Preface

Although stellar dynamics is a well established discipline with quite a long history behind it, the topic of the meeting we chose to hold in March 1993 in Geneva could have met with some surprise a few decades ago. During recent years radical changes have occurred in many areas of scientific research, changes following from the realization of the underlying universality of the concepts of chaos, randomness and unpredictability. The determinism of Laplace which once dominated physical sciences has been replaced by the image of Arnold's cat. At the same time spectacular advances in the theory of dynamical systems have provided rigorous tools to study those complex phenomena. This has gone hand in hand with the increased observational accuracy of stellar systems, which offers the opportunity to put serious empirical constraints on the theory. The idea of bringing together specialists covering these topics is initially attractive because of the possibility of obtaining the mutually useful information first hand. However, there was also an element of danger: that of having monologues instead of dialogues. But we were encouraged by the previous experience of meetings on stellar dynamics held in 1966, 1967 (Besançon, Paris). Turning now the pages of those proceedings, after more than a quarter of century, one cannot avoid the feeling of respect for the participants, how correctly they anticipated the forthcoming problems to be solved, guessed their possible answers... And if that was the time of the amazing discoveries of Hénon & Heiles, Lorenz and other pioneers, today we have the same feeling when we find regular behaviour of complex systems. This was the logical and historical origin of the idea of our Workshop.

The volume we are now presenting contains the talks given at the Workshop, with the emphasis on their review character, aiming to provide a reference book useful for as many specialists as possible from the different areas. It concerns first of all, the representation of tools of ergodic theory: along with the results of latest observations it contains sophisticated mathematical techniques usually not found in books on stellar dynamics. This is perhaps the first distinguishing feature of this volume. To maintain the emphasis on review, we therefore unfortunately had to omit the interesting discussions following almost all of the talks. However, we do present the discussion held at the last session of the Workshop concerning 10 essential unsolved problems of stellar dynamics: in theory, computer simulations and observations. This can be considered as another non-traditional aspect of the book.

Finally, let us mention that, to fulfil the conditions imposed by our aims, in accord with the Scientific Organizing Committee, all the submitted papers were refereed, almost all were revised and a few rejected, a fact again not usual for conference proceedings. Only the texts of discussion talks have been kept untouched. As a result we have a collection of papers of an even better

quality than those submitted originally. This experience may indicate to the editors and publishers that the refereeing of certain conference proceedings is not only desirable but can even be more important than for journals, if the former are intended to serve longer and for a wider audience.

Let time judge this further attempt at deciphering the truth about the world of stellar systems, for us associated perhaps with the face of another mysterious cat, the Cheshire cat of Alice in Wonderland.

We are grateful to all those who helped us in the organizing of the workshop. In particular our thanks go to the Director of Geneva Observatory, Professor André Maeder, the members of the Local Organizing Committee, Professor Louis Martinet and Dr. Daniel Friedli, and the secretaries, Ms. Irène Scheffre and Ms. Elisabeth Teichmann. We would also like to mention the help of Dr. Armen Kocharyan and the anonymous referees in the preparation of this volume.

We acknowledge the financial supports of the Swiss Academy of Natural Sciences (ASSN), the Swiss National Science Foundation (FNRS), and the University of Geneva. They have been decisive in enabling this conference to take place.

Finally, it is a particular pleasure to record the fact that the daughter of one of us, Karin Pfenniger, had the good taste to delay her birth, first announced for the beginning of the workshop, by four weeks.

Geneva
October 1993

Vahe Gurzadyan
Daniel Pfenniger

Contents

3. Computer Simulations and Mappings

4. Instabilities

5. The Few-Body Problem

6. Large-N Limit

7. Galactic and Extragalactic Problems

List of Participants

ALY Jean-Jacques
Service d'Astrophysique
CE Saclay
F-91191 Gif Sur Yvette, Cédex
FRANCE

ATHANASSOULA E.
Observatoire de Marseille
2, Place Le Verrier
F-13248 Marseille, Cédex 4
FRANCE

BALCELLS Marc
Kapteyn Astronomical Institute
Postbus 800
NL-9700 AV Groningen
THE NETHERLANDS

BIESIADA Marek
Nicolaus Copernicus Astronomical Center
Bartycka 18
PL-00-716 Warsaw
POLAND

CIPRIANI Pietro
Dipartimento di Fisica
Universita La Sapienza
Piazzale Aldo Moro 2
I-00185 Roma
ITALY

DE ZEEUW P.T.
Sterrewacht Leiden
Huygens Laboratorium
Postbus 9513
NL-2300 RA Leiden
THE NETHERLANDS

DJORGOVSKI S.
Palomar Observatory
Caltech
Pasadena CA 91125
USA

EARN David J.D.
Institute of Astronomy
Madingley Road
Cambridge CB3 OHA
UK

FRIEDLI Daniel
Geneva Observatory
University of Geneva
51, ch. des Maillettes
CH-1290 Sauverny
SWITZERLAND

FUCHS Burkhard
Astronomisches Rechen-Institut
Mönchhofstraße 12-14
D-W 6900 Heidelberg
GERMANY

FUNATO Yoko
College of Arts and Sciences
The University of Tokyo
3-8-1 Komaba
Meguro-ku, Tokyo 153
JAPAN

FUX Roger
Geneva Observatory
51, ch. des Maillettes
CH-1290 Sauverny
SWITZERLAND

GALGANI Luigi
Institute of Mathematics Milano
I-20133 Milano
ITALY

GIORGILLI Antonio
Institute of Mathematics Milano
I-20133 Milano
ITALY

GODLOWSKY Wlodzimierz
Jagiellonian University Observatory
Orla 171
30-244 Krakow
POLAND

GOUDA Naoteru
Dept. of Earth and Space Science
Faculty of Science
Osaka University
Toyonaka, Osaka 560
JAPAN

GRIV Eugeny (Uri)
Dept. of Physics
Ben-Gurion University of the Negev
P.O. Box 653
Beer-Sheva 84105
ISRAEL

GURZADYAN V.G.
Department of Theoretical Physics
Yerevan Physics Institute
Yerevan 375036
ARMENIA

HAYLI Avram
Observatoire de Lyon
Avenue Charles André
F-69561 St-Genis-Laval, Cédex
FRANCE

HEYVAERTS Jean
Observatoire de Strasbourg
11, rue de l'Université
F-67000 Strasbourg
FRANCE

INAGAKI Shogo
Dept. of Astronomy
University of Kyoto
Kyoto 606
JAPAN

KANDRUP Henry, E.
University of Florida
Dept. of Astronomy
211 Space Sciences Res. Bldg.
Gainesville FL 32611
USA

KONISHI Tetsuro
Dept. of Physics
School of Science
Nagoya University
Nagoya, 644-01
JAPAN

LASKAR Jacques
Bureau des Longitudes
77, av. Denfert-Rochereau
F-75014 Paris
FRANCE

MAHON Mary Elaine
University of Florida
Dept. of Astronomy
211 Space Sciences Res. Bldg.
Gainesville FL 32611
USA

MAKINO Junichiro
College of Arts and Sciences
University of Tokyo
3-8-1 Komaba
Meguro-ku, Tokyo 153
JAPAN

MARTINET Louis
Geneva Observatory
51, ch. des Maillettes
CH-1290 Sauverny
SWITZERLAND

MEYLAN Geo
European Southern Observatory
Karl-Schwarzschild Straße 2
D-W 8046 Garching bei München
GERMANY

MILLER R.H.
Astronomy and Astrophysics Dept.
University of Chicago
5640 Ellis Avenue
Chicago IL 60637
USA

MUZZIO Juan C.
Observatorio Astronomico
Paseo del Bosque
1900 La Plata
ARGENTINA

NEZHINSKIJ Evgenij
Sterrewacht Leiden
Huygens Laboratorium
Postbus 9513
NL-2300 RA Leiden
THE NETHERLANDS

NUNEZ J.A.
Observatorio Astronomico
Paseo del Bosque,
1900 LA PLATA
ARGENTINA

OSSIPKOV Leonid
Astronomical Institute
St. Petersburg University
Bibliotechnaya pl. 2
198904 St Petersburg
RUSSIA

PAPAPHILIPPOU Yannis
Bureau des Longitudes
77, av. Denfert-Rochereau
F-75014 Paris
FRANCE

PATSIS Panos
ESO
Karl-Schwarzschild Straße 2,
D-W 8046 Garching bei München
GERMANY

PEREZ Jérôme
Service d'Astrophysique
CE Saclay
F-91191 Gif Sur Yvette, Cédex
FRANCE

PETROVSKAYA Irina
Astronomical Institute
St. Petersburg University
Bibliotechnaya pl. 2
198904 St Petersburg
RUSSIA

PETTINI Marco
Osservatoria Astrofisico di Arcetri
Largo E. Fermi 5
I-50125 Firenze
ITALY

PFENNIGER Daniel
Geneva Observatory
University of Geneva
51, ch. des Maillettes
CH-1290 Sauverny
SWITZERLAND

PUCACCO Giuseppe
Gruppo 69 - ICRA
Universita "La Sapienza"
Piazzale Aldo Moro 2
I-00185 Roma
ITALY

QUINLAN Gerald
Lick Observatory
University of California
Santa Cruz CA 95064
USA

ROMEO Alessandro B.
Onsala Space Observatory
Chalmers University of Technology
S-43992 Onsala
SWEDEN

ROUET Jean-Louis
PMMS/CNRS
3A, av. de la Recherche Scientifique
F-45071 Orléans
FRANCE

SMITH Haywood
Department of Astronomy
University of Florida
211 Space Sciences Res. Bldg.
Gainsville FL 32611
USA

SPURZEM Rainer
Institute for Theoretical Physics
and Observatory
University of Kiel
Olshausenstraße 40
D-W 2300 Kiel
GERMANY

SZYDLOWSKI Marek
Jagiellonian University Observatory
Orla 171
30-244 Krakow
POLAND

TABIRYAN Nelson
Institute für Angewandte Physik
Hochschulstraße 6
D-W 6100 Darmstadt
GERMANY

TSUCHIYA Toshio
Faculty of Science
Kyoto University
Kitashirakawa-oiwakecho Sakyoku
Kyoto
JAPAN

VAN ALBADA Tjeerd S.
Kapteyn Astronomical Institute
Postbus 800
NL-9700 AV Groningen
THE NETHERLANDS

WOZNIAK Hervé
Geneva Observatoty
University of Geneva
51, ch. des Maillettes
CH-1290 Sauverny
SWITZERLAND

Foreword

Future historians of astronomy will certainly recognize as a major step in our field, the progress enabling us to precisely reconstruct the astrophysical objects by numerical simulations on computers. This provides an iterative process, with a mutual feedback between theory and observations. This is an essential procedure in the progress of understanding the deep physics and evolution of planets, stars, galaxies and the Universe. The present Workshop on "Ergodic Concepts in Stellar Dynamics" organized by V.G. Gurdzadyan and D. Pfenniger is a very valuable contribution to this procedure of understanding star clusters and galaxies.

Stellar dynamics is basically governed by a simple law: Newton's law of gravitation. It is thus amazing, at least for the nonspecialist, how different physical phenomena may arise in this context: from Keplerian orbits to chaotic motion, instabilities and gravothermal collapse. The basic law is simple, but the consequences are incredibly rich, as usual in Nature.

The unavoidable counterpart of this natural complexity, which gives some additional flavour to the life of astrophysicists, is that the handling of the problems requires special mathematical and numerical tools, as well as extreme care and astuteness in the work and interpretation. A major concern of this workshop and of the associated volume is to present the leading contributions in this field with a well balanced distribution between the various aspects of the problem: basic observations, mathematical methods, numerical tools, instabilities both physical and numerical, systems with large numbers of bodies. To provide the icing on the cake there is also a nice final discussion.

Geneva Observatory was founded in 1772. It has a long tradition of quality in astronomy, with very active groups in various fields, such as stellar dynamics and kinematics, dynamics and evolution of galaxies, stellar physics and evolution, stellar photometry, high energy astrophysics and active galactic nuclei. This workshop is a highlight among these activities and I may express the deep gratitude of Geneva Observatory and of the scientists working in this field to V.G. Gurzadyan and D. Pfenniger for having taken the initiative in organizing this meeting and for having conducted it to a well-deserved success.

André Maeder
Director of Geneva Observatory

1. Observations

Structural and Dynamical Forms of Elliptical and Dwarf Galaxies

S. DJORGOVSKI

Palomar Observatory, 105–24 Caltech, Pasadena, CA 91125, USA

Abstract. This review discusses two intriguing problems posed by observations of early-type galaxies: First, the existence of a so-called fundamental plane of elliptical galaxies and its thickness impose strong constraints on their global dynamical structures, stellar populations, and the relation of dark to visible matter in them. Second, the prevalence of extreme core-halo structures among the nucleated dwarf galaxies. Finally, some general remarks are made about the standard density forms of galactic subsystems, including dark halos. These empirical findings provide strong constraints and valuable clues for models of galaxy formation. They may point towards the existence of preferred dynamical structures in galaxies, perhaps globally stable or maximum entropy states, reached through different formative histories.

1. Introduction: Galaxy Families and Galaxy Manifolds

Understanding of galaxies, their structure and physics is inseparable from the understanding of their formation and evolution. Silk & Wyse (1993) give a good recent review of this subject, in its many aspects. While searches for young galaxies at large redshifts aim to probe directly their early evolutionary phases, as far as the formation of *normal* galaxies is concerned, such studies have met with only a limited success so far, due to the inherent difficulties in observations of faint objects; for a recent review, cf. Djorgovski (1992c), and references therein.

An alternative path is in systematic studies of relatively nearby galaxies, which offer a richer signal and can be studied in much more detail. In addition to the understanding of galaxies themselves, such studies can provide interesting problems and insights for the dynamics of stellar systems. As global systematics of galaxian properties begin to emerge, one can try to make inferences about the possible formative mechanisms and evolutionary paths which could have produced the observed phenomenology of galaxies today. There has been a substantial progress in this area over the past few years.

Traditionally, systematic studies of galaxy properties tended to be within a loose framework which may be called the Hubble sequence paradigm. This framework has served us well, but its inadequacies are increasingly becoming apparent. Vast amounts of new, quantitative information neither fit easily into the old morphological bins nor can be explained within such purely qualitative, empirical schemes. For example, the Hubble sequence fails completely in describing the properties of elliptical galaxies and correlations among them; the fundamental dichotomy between dwarf and normal (giant) galaxies has been completely missed until a decade ago; and so on. Galaxy families should be defined quantitatively, by the portions of the parameter space they occupy and the correlations they obey, rather by the superficialities of visible light morphology.

A new operational framework is needed for the modern, quantitative, systematic studies of galaxy properties, their physics and evolution; perhaps a galaxian equivalent to the Hertzsprung-Russel diagram as a tool for studies of the physics and evolution of stars. A schematic concept of such a framework, called galaxy parameter space, has been proposed by Djorgovski (1992a, 1992b).

Both ellipticals and spirals (using the word "spirals" as a shorthand for galaxies with gaseous, star-forming disks, regardless of the density wave pattern therein) form statistically two-dimensional sequences or manifolds in the parameter spaces of many of their observed global properties. In other words, a multitude of independent global physical properties which can be defined and measured, such as the luminosity, characteristic radius, mass, various forms of density, characteristic velocity scales, colors and other stellar population variables, etc., can be fully described to within the measurement errors by only two numbers. For elliptical galaxies, this manifold has been named the Fundamental Plane (hereafter FP), and is described in the following section. For spirals, where different observables are used, there is a so-called Scale-Form plane (Whitmore 1984); the Tully-Fisher relation represents a nearly edge-on view of this manifold. For both galaxy families, many global properties are connected by sets of excellent bivariate correlations. These correlations must be products of galaxy formation and evolution, and as such represent precisely the empirical clues for understanding of galaxies we have been looking for.

Pursuing the analogy with the H-R diagram, just as stars form one-dimensional sequences of mass embedded in a parameter space of luminosity and temperature, galaxies form two-dimensional sequences embedded in a parameter space of a "size" (mass, luminosity, or radius), density (or surface brightness), and kinetic temperature (e.g., velocity dispersion for pressure-supported systems, or rotational speed for disks). The FP is an analog of the "main sequence" for ellipticals. Just like stars follow evolutionary tracks in the H-R diagram, galaxies evolve through their parameter space in the ways which yet remain to be fully understood.

Given the complex nature and structure of galaxies, it is quite remarkable that the statistical dimensionality of these manifolds is only two. Why this might be in the case of elliptical galaxies is discussed below. In the case of spirals, the implication is that the star formation history of the disks is somehow tightly regulated by the properties of their dark halos. In either case, this reduction of dimensionality in galaxy manifolds implies that there are strong global regularities in both equilibrium density structures and dynamics of galaxies: Hamiltonian (dissipationless) systems cannot "forget" dynamical information. One cannot help but feel the invisible hand of entropy at work, setting up the basic properties of galaxies as we know them.

2. The Fundamental Plane of Elliptical Galaxies and its Implications

A number of the observable and derived physical properties of elliptical galaxies, including luminosities, masses, all consistently defined measures of radius (core, half-light, isophotal, etc.), projected velocity dispersions, all consistently defined measures of surface brightness or density, colors, and measures of metallicity (line strengths), are connected in a statistically two-dimensional manifold, the so-called Fundamental Plane (FP), by a set of bivariate scaling laws; i.e., the correlations are linear when the logarithmic quantities are used. However, a number of other quantities, generally describing the azimuthal and radial shapes and distribution of the luminous material (e.g., ellipticity, measures of triaxiality, etc.), the velocity anisotropy, or the stellar population gradients, *do not* participate in the FP, and do not correlate well with any other property. The subject is reviewed, e.g., by Djorgovski (1987, 1992a, 1992b), Kormendy & Djorgovski (1989), Djorgovski & de Carvalho (1990), Bender et al. (1993), Capaccioli et al. (1993), or Djorgovski & Santiago (1993), and references therein.

Simple or minimal representations of the FP can be made in three-dimensional spaces of observable or derived quantities. If logarithmic quantities are used, the data points sit on a plane, which is generally tilted with respect to all observable axes. The coordinate axes given by the data themselves, i.e., the eigenvectors of the data distribution, are not simply related to any observable or derived quantities, although the principal eigenvector is close to the galaxy mass axis (Djorgovski 1992b). The coordinate system can be rotated or tilted to suit a given purpose. One particular possibility among the infinitely many was proposed, e.g., by Bender et al. (1992), who renamed the mass to κ_1, and mass-to-light ratio (M/L) to κ_3. However, their $\{\kappa_1; \kappa_2\}$ plane is tilted with respect to the FP, i.e., $\kappa_3 \neq$ const., so the advantages of this particular choice of a coordinate system are not clear. It is perhaps cleaner to stick with the directly observable quantities.

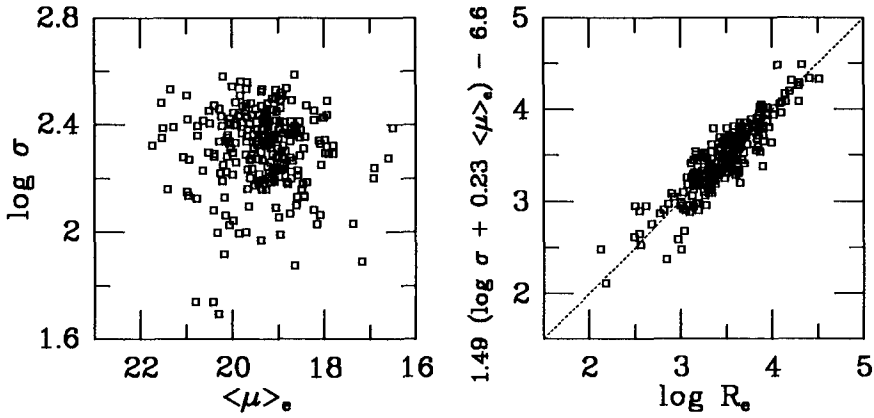


Fig. 1. Parameters of elliptical galaxies, from a survey described by Penereiro et al. (1993). The left panel shows a plot of the projected central velocity dispersion σ, measured in km/s, versus the mean surface brightness within the effective (half-light) radius isophote $\langle\mu\rangle_e$, measured in red magnitudes per arcsec^2. This is an observer's equivalent of the cooling diagram from theories of galaxy formation. It represents a view of the FP nearly face-on: the observed scatter far exceeds the measurement errors. The right panel shows a bivariate correlation between a combination of velocity dispersion and surface brightness, and the effective radius R_e, measured in parsecs. This is a view of the FP edge-on: the residual scatter is fully accounted by the measurement errors. The two views illustrate the statistical two-dimensionality of the manifold of elliptical galaxies.

A common practice is to consider the parameter space whose axes include a measure of radius or semimajor axis, R, such as the de Vaucouleurs' effective radius (but other possibilities would do as well), projected central velocity dispersion, σ, and the mean surface brightness within that radius or the corresponding elliptical isophote (using notation I for the projected surface brightness in linear units, such as the solar luminosities per unit area, or μ for the corresponding logarithmic quantity in magnitudes per square arcsec[1]). In this notation, the common expression of the FP is a bivariate scaling relation:

$$R \sim \sigma^A \, I^B \, , \tag{1}$$

where the typical observed values of the coefficients are $A \simeq 1.4$ and $B \simeq -0.85$, with uncertainties of the order of 10%. There are also comparable variations between different data samples, and dependences on the photometric bandpass and the large-scale environment (de Carvalho & Djorgovski 1992a; Djorgovski & Santiago 1993). Figure 1 shows a real-life example from the data set on ellipticals described by Penereiro et al. (1993).

[1] The rational reader is rightly offended by such astronomical arcana, but alas, these units are commonly used and are hard to avoid.

A heuristic derivation of the FP was presented by Djorgovski et al. (1988). Basically, for a family of objects bound by Newtonian gravity (or any r^{-2} force), the virial theorem (hereafter VT) produces the following scaling relation:

$$\langle R \rangle \sim \langle V^2 \rangle \, \langle I \rangle^{-1} (M/L)^{-1} \, , \tag{2}$$

where $\langle R \rangle$ is the mean mass radius, $\langle V^2 \rangle$ is the mean kinetic energy per unit mass, i.e., a velocity scale, $\langle I \rangle$ is the mean surface brightness, defined so that the luminosity is proportional to $\langle I \rangle \, \langle R \rangle^2$, and (M/L) is the mass-to-light ratio. These quantities may be related to the observed ones by some arbitrary functions. The simplest case is that the corresponding quantities are roughly proportional, e.g., $\langle R \rangle = k_R R$, $\langle V^2 \rangle = k_V \sigma^2$, $\langle I \rangle = k_I I$, etc., where the proportionality coefficients themselves may be functions of other parameters, e.g., the total mass, velocity anisotropy, and so on. These coefficients reflect the dynamical and density structure of the galaxies. These proportionalities could also be, in principle, intrinsically very noisy.

The VT connects three global parameters by a single equation. Thus, it represents an equation of a plane in a parameter space of any three quantities connected by it, such as the mass M, radius $\langle R \rangle$, velocity scale, or various forms of volume density or projected (surface) density. If galaxies also sit on a plane in the parameter space of the observed quantities, then the following condition must apply:

$$k_V^{-1} \, k_I \, k_R \, (M/L) \sim \sigma^{A-2} \, I^{-B-1} \, . \tag{3}$$

Note that the various k's need not be constants, but functions of any number of galaxian properties, reflecting their internal dynamics, projection effects, etc. The (M/L) ratios depend on the ratio of visible to dark matter in galaxies, and their spatial distribution (mainly, the relative amounts of the dark matter within the visible portions of galaxies); they also reflect the stellar populations of galaxies, through the relative numbers of low-mass dwarfs, white dwarfs, or dark stellar remnants, which contribute little light, but could lock up substantial portion of the baryonic mass.

The mapping of the VT plane into the observed FP, illustrated schematically in Fig. 2, requires that the nontrivial condition given by Eq. (3) is satisfied, and thus reflects some deep regularities in the composition and dynamics of elliptical galaxies. First, the small observed thickness of the FP implies that there can be little or no intrinsic scatter in Eq. (3): either all k's and (M/L)'s are constants or weak functions of the FP variables with little or no scatter in each of them, or their scatter has to be strongly correlated and mutually compensated. Neither possibility is easy to arrange. Second, the observed values of the FP slope coefficients are $A \neq 2$, and $B \neq -1$, and thus *something* has to be changing systematically within the FP, in just the right amount to produce the observed values of these coefficients, and yet not introduce a scatter which would thicken the FP at any given location. Both

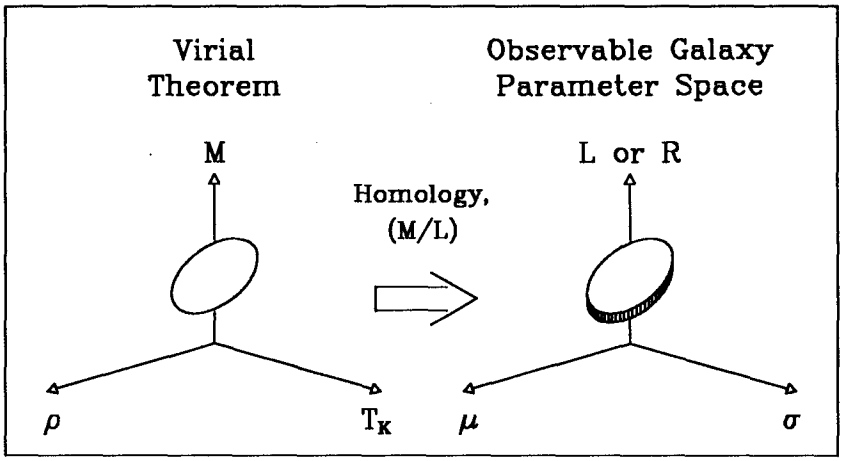


Fig. 2. A schematic illustration of the problem posed by the FP; all axes are assumed to represent logarithms of the indicated quantities. Galaxies are in general guaranteed to sit on a plane defined by the virial theorem (VT), in the parameter space of their mass M, mean density ρ, and mean kinetic temperature (or a characteristic velocity squared) T_K. In the parameter space of observable quantities, such as the luminosity L or a radius R, surface brightness μ, and projected velocity dispersion σ, elliptical galaxies form a two-dimensional sequence, the FP. The observed thickness of the FP is comparable to the measurement errors, although there might be some intrinsic thickening as well. The mapping of the VT plane into the observed FP requires that non-trivial constraints about homology of density and dynamical structures of ellipticals, and their mass-to-light ratios are satisfied

of these implications of the FP thus impose strong constraints on formation and dynamics of elliptical galaxies.

There are two basic approaches to the problem: through dynamics (k_R, k_V, k_I), and through (M/L) ratios; of course, they are not mutually exclusive. If the observed tilt of the FP with respect to the pure-VT values $A = 2$ and $B = -1$ is blamed entirely on the (M/L) ratios, the implied scaling law is $(M/L) \sim M^{\alpha}$, where α is typically $\sim 0.2 - 0.4$, depending on the sample, bandpass, etc. (Faber et al. 1987; Djorgovski 1988; Djorgovski & Santiago 1993).

It is generally believed that most of the mass within the visible parts of ellipticals is baryonic. Furthermore, measures of the average stellar metallicity are also contained in the FP (Djorgovski 1987; de Carvalho & Djorgovski 1989). It is thus tempting to ascribe the variations in the (M/L) to differences in the initial mass function of stars (Djorgovski 1988). Renzini & Ciotti (1993) have explored this possibility in some detail, and concluded that a considerable fine-tuning would be needed in order to reproduce both the tilt of the FP, and the small scatter around it. It seems more likely that the (M/L) of stellar populations (i.e., excluding the dark matter), and thus the initial mass functions of stars are universal and constant for most (old) ellipticals. This

is in itself a non-trivial cosmological observation which should be explained, and it also leaves us with the FP problem.

A variation of the dark to luminous matter ratio, or in their relative distributions could be invoked, in the sense that more luminous or more massive galaxies have fewer baryons and/or fewer stars produced per unit mass. As there is no believable and complete theory of dissipative galaxy formation yet, little can be said about this possibility. Reproducing the observed scaling of the apparent (M/L) with mass would be a good test for the models of galaxy formation. It is intriguing that such a relation was indeed found by Navarro (1993), but obviously the case is not closed yet. Even if there is such a scaling relation, the scatter should still be relatively small, implying a remarkable constancy in the ratio of dark to luminous mass at any given point in the FP, or even globally, for the entire family of ellipticals.

Regardless of whether the variation in (M/L)'s could explain the tilt of the FP, we are still left with a near constancy of the structural coefficients k_R, k_V, and k_I. If all elliptical galaxies had exactly the same density structure, and exactly the same dynamics, these coefficients would be constant. However, we know that ellipticals show a very considerable diversity of stellar kinematics, and a variety of shapes and density profiles. In addition to the frequent occurrence of "peculiar" kinematics, such as the counter-rotating cores, minor axis rotation, etc., ellipticals show a broad range of velocity anisotropy or the relative importance of the rotational to pressure support, as parametrized, e.g., by the ratio of the maximum rotational speed to the projected velocity dispersion $(V_m/\sigma)^*$ (Davies et al. 1983; see Kormendy & Djorgovski 1989, or de Zeeuw & Franx 1991 for reviews). Busarello et al. (1992) find no correlation between the rotational and random components of the total kinetic energy. Furthermore, none of the observed measures of velocity anisotropy, or of the radial or azimuthal shapes of light distribution correlate with any of the FP variables, and all of them show a considerable scatter. Even if one divides the samples of ellipticals by the median $(V_m/\sigma)^*$, ellipticity, or the boxiness/diskiness parameter a_4, which is another measure of velocity anisotropy, one obtains the FP solutions only marginally different (Djorgovski & Santiago 1993).

Therein lies the paradox: how is it possible to have such a great observed diversity of shapes and kinematics, yet have the the structural and dynamical properties coupled so rigidly so that there is essentially no intrinsic scatter at any point within the FP? If one accounts for the measurement errors, the thickness of the FP in the velocity dispersion direction corresponds to the scatter of at most a few percent in k_V (Djorgovski & Santiago 1993). It would be very easy to generate a scatter much larger (of the order of 100%!) by relatively modest amounts of velocity anisotropy (Tonry 1983; Merritt 1988). Why are ellipticals so well standardized?

While the observed surface brightness (and thus density) profiles of ellipticals do show some variety, they are broadly similar, following the general

Hubble – de Vaucouleurs shape: they can be reasonably well described either by a slightly curved power-law profile with a slope $\simeq -2$ and a core, or as the $r^{1/4}$ profile with some deviations (cf. Kormendy 1982; Djorgovski 1985; Capaccioli 1988; Kormendy & Djorgovski 1989, and references therein). There is a weak and noisy trend in that more luminous ellipticals tend to have shallower profiles (Schombert 1987), but there is a variety of profile shapes at every luminosity. Nevertheless, the general similarity of profiles does suggest that k_R may be close to being a constant, or only a weak function of the galaxy mass.

There is some incipient understanding as to why ellipticals have such density distributions. Objects with density profiles with that general shape are readily produced in numerical simulations of cold collapse (van Albada 1982; Villumsen 1984; McGlynn 1984; Aguilar 1988; Aguilar & Merritt 1990; and many others). Adding dissipation does not seem to change the basic outcome (Carlberg et al. 1986); this is important, since the formation of ellipticals must have involved a considerable degree of dissipation (cf. Kormendy & Djorgovski 1989 for detailed arguments). The relative kinetic temperature, collapse factor, presence or absence of a secondary infall, preexistence and the radial density distribution of a dark halo, and the relative amount of a net rotation all play some role in determining the details of the resulting structure. Variation in such parameters could easily explain the observed diversity of profile shapes. However, these are just *simulations*, and do not answer the basic question: why is there a standard product, a density profile of the Hubble – de Vaucouleurs type?

The answer almost certainly has to involve maximization of entropy in some form, achieved through violent relaxation (Lynden-Bell 1967), or some other type of collective relaxation process (e.g., Gurzadyan & Savvidy 1986; or Pfenniger 1986). Functional forms of the phase space density distributions which both seem to reproduce the observed density profiles of ellipticals, and maximize some entropy-like functional, have been proposed by Binney (1982), Bertin & Stiavelli (1984), Stiavelli & Bertin (1985, 1987), Tremaine (1987), Maoz & Bekenstein (1990), Ciotti (1991), Spergel & Hernquist (1992), and others; for recent reviews, see Binney (1988), or Bertin & Stiavelli (1993). Still, physical circumstances which would lead to a "natural selection" of a particular distribution remain poorly understood. It is possible that violent relaxation or some similar process was at work, since formation of ellipticals probably involved a good deal of merging, regardless of the amount of dissipation involved; dissipation in general also tends to erase information. This is an outstanding challenge for the theory.

The uniformity of (M/L) ratios and density distributions is not enough to guarantee the existence of a thin FP; uniformity of their dynamical structures is also required, as demonstrated by the strong observational constraints on k_V. One way to express this is that given the observed values of the galaxy's radius and mean surface brightness (both defined in some consistent manner,

but still with a considerable algorithmic latitude), the observed projected velocity dispersion can be predicted to within the measurement errors! This is true almost regardless of the value of the velocity anisotropy parameters such as $(V_m/\sigma)^*$, and the details of kinematics; the projection effects also seem to be neatly compensated. This robustness is truly remarkable: it implies a very narrow range of the net velocity anisotropy at any given location in the FP. In other words, elliptical galaxies apparently occupy only a small subset of dynamical or orbital structures which are allowed to them. A spontaneous self-selection of a restricted range of orbital structures has been seen in numerical simulations by Aguilar & Velázquez (1993).

Franx et al. (1991) found that ellipticals show a limited range of misalignment angles between their rotational axes and projected minor axes. If most ellipticals are triaxial, and the distribution of their angular momenta was uniform, a much broader distribution of kinematical misalignment angles would be expected. This again points towards a restricted set of dynamical structures.

The observed small thickness of the FP requires that ellipticals follow a narrowly defined range of dynamical and density structures. The similarity of their density profiles follows directly from simple surface photometry. The similarity of dynamical structures is far less obvious in the direct kinematical data, but it is strongly implied by the small thickness of the FP. Whereas there have been some attempts to explain the radial density structure of ellipticals, the uniformity of their dynamical structures comes as a surprise, and is almost completely unexplored. Whatever physical processes select the preferred density profiles, probably select a correspondingly narrow range of dynamical structures. Understanding of this natural selection is the principal theoretical puzzle posed by the FP.

Unfortunately, gravothermodynamical entropy is an elusive concept. For an unconfined self-gravitating system, there is no maximum entropy state (Tremaine et al. 1986). The entropy increases as a core-halo structure is established, with a compact core which contains a small fraction of the total mass, but has most of the total binding energy, and an extended, but loosely bound halo containing most of the mass. White & Narayan (1987) obtained just such solutions for isolated, self-gravitating systems of a finite mass and energy. Imposing a maximum phase space density limit simply limits the relative size of the compact core, but the basic core-halo structure remains. Only if an additional restriction was imposed that the profile should be a power-law, a Hubble-like profile shape was recovered. White & Narayan concluded that some additional, as yet unknown physical constraints must be operating during the formation of ellipticals (or, for that matter, in numerical simulations), in order to reproduce the observed density profiles.

3. Extreme Core-Halo Structures in Dwarf Galaxies

As it happens, the core-halo structures *do* occur in vast numbers of galaxies, namely the low surface brightness dwarfs (LSBD). The core-halo contrasts can be rather extreme, and the apparent similarity with the maximum-entropy solutions is enticing. What are these objects and how did they achieve such states?

LSBD are the most numerous galaxies known in the local universe, even if they do not contribute most of the total light (Sandage et al. 1985). Historically they have been confused as "dwarf ellipticals". They certainly have nothing to do with the true ellipticals, as their properties and correlations between them are vastly different (Wirth & Gallagher 1984; Kormendy 1985; Kormendy & Djorgovski 1989; de Carvalho & Djorgovski 1992b; Djorgovski 1993, and references therein). Not all of them are nucleated, but that may be only a detection threshold effect: the nuclei are unresolved or barely resolved even in the best ground-based observations. Many are, and the frequency of occurrence and the fraction of light contained in the compact nucleus both tend to increase with the increasing luminosity of the host galaxy; typically, the nucleus contains a few percent of the total galaxy light. The most extensive study of their photometric properties to date is that by Binggeli & Cameron (1991).

The nearest case to us is NGC 205, a dwarf companion of the Andromeda galaxy, M31. Unfortunately, it is a relatively unimpressive case of a compact nucleus, relative to the typical cases studied in the Virgo and Fornax clusters and elsewhere by Binggeli and his collaborators, and others. Its surface brightness profile is shown in Fig. 3, along with an example of a normal elliptical of a comparable luminosity, at a comparable distance. The logarithmic representation is really hiding the contrast, and a linear intensity cut through NGC 205 is shown in Fig. 4. Even the deconvolved version of the ground-based data underestimates the true contrast, as shown by the Hubble Space Telescope observations (I.R. King, priv. comm.).

The dynamics of NGC 205 has been measured by Held et al. (1990) and by Carter & Sadler (1990). Both studies indicate that the nucleus is *colder* than the surrounding galaxy, with the projected velocity dispersions $\sigma_{\rm NUC} \simeq 20$ km/s, and $\sigma_{\rm GAL} \simeq 60$ km/s.

Both groups conclude that NGC 205 and presumably other LSB dwarfs are most likely supported by velocity anisotropy, rather than by rotation (cf. also Bender et al. 1991).

The derived properties of the nucleus and the galaxy from the seeing-deconvolved ground-based observations (Bendinelli et al. , in prep.) are listed in Table 1. The contrast between the nucleus and the galaxy is rather striking: the nucleus is at least 4 orders of magnitude denser than the surrounding galaxy, and it is thus clearly a dynamically independent system. We should also bear in mind that this is a relatively puny example of a compact nucleus,

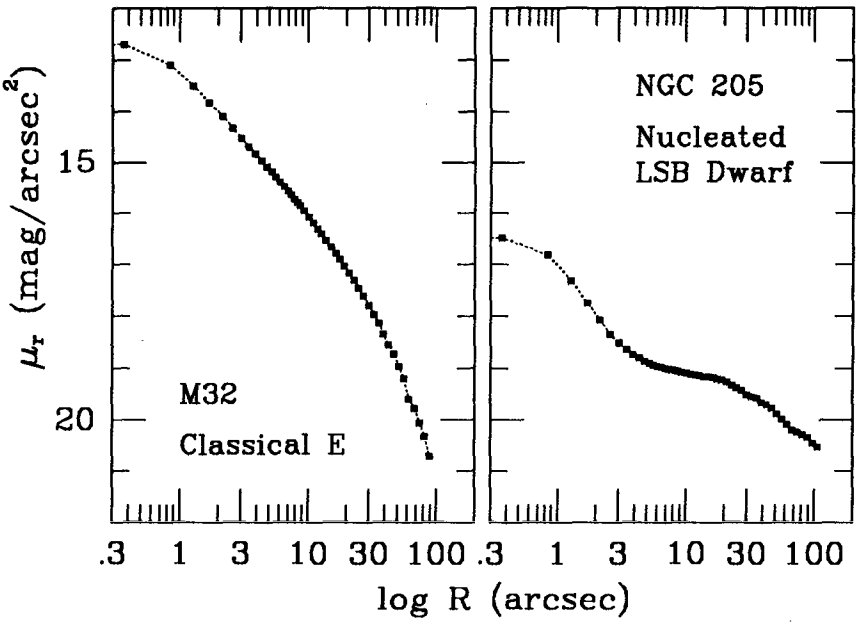


Fig. 3. Surface brightness profiles of a normal or "classical" elliptical galaxy M32 = NGC 221 (left), and a nucleated low surface brightness dwarf galaxy, NGC 205 (right), plotted on the same scale. The data were taken from a uniform survey by Djorgovski (1985). The two objects have comparable luminosities and distances, and illustrate the contrast between the true ellipticals and dwarf galaxy families. The nucleus of NGC 205 is barely resolved, and its prominence is clearly limited by the seeing. Logarithmic representation hides the real contrast, as illustrated in Fig. 4

Table 1. Physical parameters of the NGC 205 nucleus and its host galaxy

	Nucleus (deconvolved)	Galaxy
Core radius	$\lesssim 0.3'' \simeq 1.0$ pc	$25'' \simeq 87$ pc
Half-light radius	$\lesssim 0.4'' \simeq 1.4$ pc	$100'' \simeq 350$ pc
Luminosity	$5.5 \times 10^5\ L_\odot$	$3.7 \times 10^8\ L_\odot$
Central luminosity density	$\gtrsim 3.6 \times 10^4\ L_\odot/\mathrm{pc}^3$	$4.6\ L_\odot/\mathrm{pc}^3$
Average luminosity density	$\gtrsim 1.6 \times 10^4\ L_\odot/\mathrm{pc}^3$	$0.7\ L_\odot/\mathrm{pc}^3$
Central relaxation time	$\lesssim 8 \times 10^7$ yr	10^{12} yr
Mass-to-light ratio	$0.9\ M_\odot/L_{\odot,\mathrm{V}}$	$17\ M_\odot/L_{\odot,\mathrm{V}}$

and that these data are still limited by the seeing; for most nucleated dwarfs, the contrasts are probably even stronger.

The origin of these systems is still unknown. Both photometry and spectroscopy indicate that the nuclei are perhaps *slightly* younger than their host

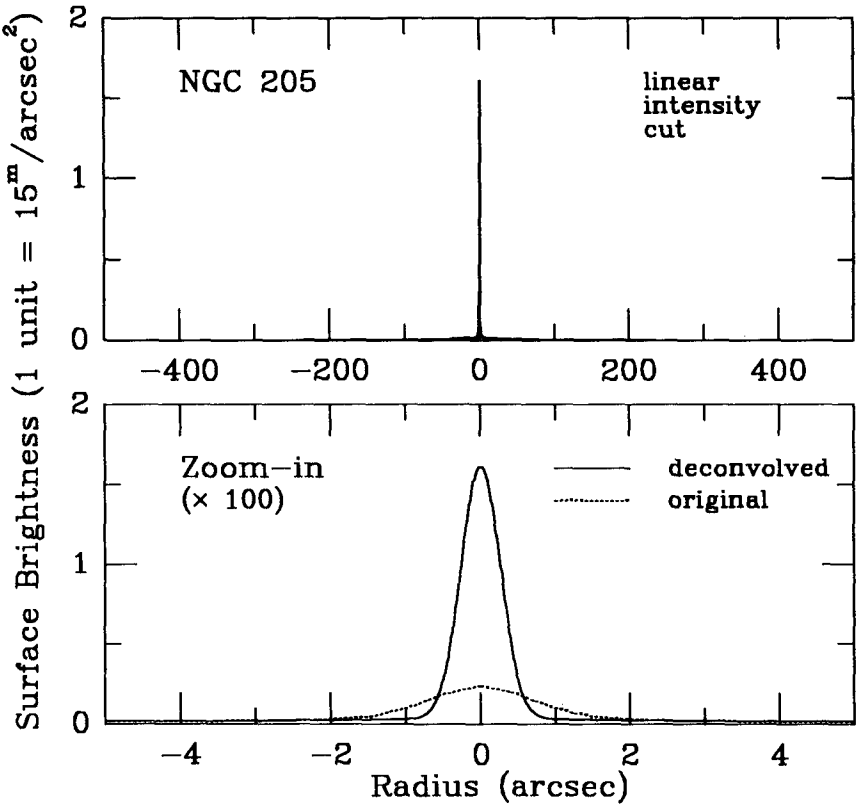


Fig. 4. Linear intensity cuts through the center of NGC 205 illustrate the dramatic core-halo nature of its surface brightness distribution. The lower panel shows a zoomed-in section of the cut, both before and after the partial seeing deconvolution (Djorgovski et al. 1992; Bendinelli et al., in preparation). This deconvolved profile is still seeing-limited; unpublished HST images by Ivan King (priv. comm.) show an even higher intensity contrast. It should be kept in mind that NGC 205 nucleus is a relatively inconspicuous representative of the class!

galaxies (Vigroux et al. 1984; Bothun & Mould 1988; Vader et al. 1988; etc.). They could have formed from delayed bursts of star formation, perhaps from the gas released by the normal stellar evolution, or accreted from the outside, and accumulated at galaxy centers. Nuclei may also represent piles of debris of globular clusters and/or remnants of off-nuclear star forming regions which spiraled in due to dynamical friction, as in the models by Tremaine (1976). Whatever their exact formation mechanism might be, they represent fascinating dynamical systems, the closest realizations of unrestricted maximum-entropy states we know of in the world of galaxies.

4. Concluding Remarks: The Standard Forms of Galactic Subsystems, and Their Origins

With the hindsight of the knowledge gained over the past half-century, we can generalize the concept of stellar populations introduced by Baade to something more encompassing. Galaxies appear to consist of *subsystems*, where stars may be the most easily visible, but not necessarily the dominant component. These subsystems are characterized by their density distributions, dynamics, composition (stars, gas, dark matter), physical parameters of the components (e.g., average ages of stars, thermodynamical state and phase structure of the ISM, etc.), locus within the galaxy, etc. The number of qualitatively different types of subsystems is limited: for example, in our Galaxy we can distinguish the thin disk, the thick disk, the metal-rich bulge, the metal-poor stellar halo, and the dark halo; at most a few other types can be distinguished in other galaxies, for example the x-ray halos of ellipticals. Normal galaxies are composed of such building blocks in varying proportions, and this is what defines their morphological classification. Obviously, there are no sharp boundaries between them, and they are nested within the same potential wells.

Distinct properties of galactic subsystems suggest that they have formed through distinct processes and/or at distinct phases of galaxy formation. By trying to interpret their possible origins from their systematic properties, we may be able to map out the formative history of galaxies. For example, there appear to be only a few types of density distributions, as listed in Table 2. They are ubiquitous, perhaps representing "attractors" in the space of all possible density distributions – an idea which may be worth exploring.

Table 2. Characteristic density structures of galactic subsystems

Systems	Density distribution	Possible formation mech.
Ellipticals and n-body models	Hubble–de Vaucouleurs type profiles	Merging, rapid relaxation, with or w/o dissipation
Dwarfs	Core-halo structures, exponential ellipsoids?	Galactic winds and adiabatic expansion
Disks of spirals	Bi-exponential disks	Spinup, gradual infall
Dark halos	Isothermal spheres with finite cores	Early collapse, merging, gravothermal instability?

In the preceding sections we discussed the Hubble – de Vaucouleurs profiles of ellipticals and the core-halo structure of dwarfs. The former seem to be reasonably well described as restricted maximum entropy solutions, but the origin of the power-law constraint remains unclear. The later appear to be *prima facie* examples of unrestricted maximum entropy solutions. It is

likely that the formation of ellipticals (as well as the products of cold collapse n-body simulations) involved a good deal of merging, regardless of the dissipation and star formation. On the other hand, dwarfs can hardly be merger products! Most models of their formation involve adiabatic mass loss and expansion, driven by a galactic wind (Dekel & Silk 1986; Vader 1986; Yoshii & Arimoto 1987; etc.). The role of dark halos in determining the dominant evolutionary paths and the final density configurations of the visible material is not yet well understood, but obviously it could be very important, given that they account for $\gtrsim 90\%$ of the total mass (cf. Burkert 1993).

Formation of the (nearly) exponential disks is almost certainly accomplished by an extended infall within tidally spun-up dark halos. As shown by Mestel (1963), if the angular momentum is conserved and initially uniformly distributed, the resulting mass distribution would closely resemble an exponential disk, with an outer edge determined by the maximum angular momentum material. This is an intriguing, but by no means a definitive explanation. For example, Pfenniger (1989) proposed an appealing alternative. There might well be a deep dynamical reason why such structures are commonly generated in rotationally supported systems, and the role of the massive dark halos within which they reside may be critical. It is intriguing that surface brightness profiles of dwarf galaxies are also close to being exponential, but the data so far indicate that these systems are supported by velocity anisotropy, rather than by rotation.

Density distributions within dark halos are readily deduced from the flat rotation curves of spirals, i.e., $\rho(r) \sim r^{-2}$. Presumably the dark halos have cores, since the rotation curves drop as $r \rightarrow 0$. This is a density structure ($\sim$ an isothermal sphere) which is naturally generated as a product of the gravothermal instability, or the core collapse (Hénon 1961). Such structures are now readily observed in many Galactic globular clusters (Djorgovski & King 1986), where the relaxation times are short enough for core collapse to occur within the Hubble time. In fact, core collapse is the only process which we know to actually produce such a density profile in the real universe. But how can massive dark halos, which presumably originated from random ripples in the primordial density field, be so highly dynamically evolved?

One way of shortening their relaxation times is to postulate that they consist, or did consist, of very massive particles, perhaps $M_{\rm dm} \gtrsim 10^6\,M_\odot$, since for both Spitzer-Chandrasekhar two-body relaxation time, and Gurzadyan-Savvidy collective relaxation, $t_r \sim m^{-1}$. Lacey & Ostriker (1985) proposed that the dark matter might consist of $\sim 10^6\,M_\odot$ black holes, but this hypothesis may be in conflict with the data on wide binaries (Bahcall et al. 1985). Suppose, on the other hand, that the dark matter consisted of compact star clusters, possibly made of brown dwarfs (Ashman & Carr 1988) of such a mass, which is indeed close to the Jeans' mass at the recombination, $M_J \sim 10^6 M_\odot$ (Peebles & Dicke 1968). Then perhaps the dark halos could have been formed as aggregates of such primordial clusters. Increasing

the mass of the primordial halo "particles" would shorten their two-body relaxation times appreciably. Furthermore, given the inevitable clumpiness of these proto-halos, appealing to faster relaxation processes, such as the violent relaxation or the Gurzadyan & Savvidy collective relaxation, could well have lead to a rapid, early core collapse of dark halos. In this way one would naturally produce the r^{-2} density profiles. The hypothetical building-blocks dark clusters might be by now completely disrupted due to merging and tidal effects, thus avoiding any conflict with the statistics of binary separations. Their remnants might be observable as halo brown dwarfs, either individually, e.g., through direct imaging searches or gravitational microlensing, or collectively, e.g., through their IR emission (cf. Daly & McLaughlin 1992).

While a number of distinct density forms can be defined among galactic subsystems, the dynamical or phase space density forms are trickier to disentangle. The dominant mass component (usually the dark halo) determines the amplitude of the velocity field needed for support. Not much can be said beyond the trivial statement that disks are mostly supported by rotation, and other systems mostly by pressure. However, as the thickness of the FP tells us, some selection of preferred dynamical structures does occur, at least among the ellipticals. One can speculate that maximum entropy or stable configurations among self-gravitating systems will select not only specific types of density distributions, but also specific types of *phase space* density distributions.

All this is just a very tentative, broad-brush picture. However, these rather general observations do point out a number of regularities in the structures of galaxies and galactic subsystems. Their understanding is a major challenge for the theory, and will hopefully lead to new insights in the physics of collective phenomena in self-gravitating systems.

Acknowledgements. The author wishes to thank to the organizers of the conference for their hospitality and the saintly patience while waiting for this written contribution. This work was supported in part by the NASA AISRP contract NAS5-31348 and the NSF PYI award AST-9157412.

References

Aguilar L., 1898, Cel. Mech. 41, 3
Aguilar L., Merritt D., 1990, ApJ 354, 33
Aguilar L., Velázquez M., 1993, in The ESO/EIPC Workshop on Structure, Dynamics, and Chemical Evolution of Early-Type Galaxies, eds. J. Danziger et al., ESO publication No. 45, 347
Ashman K., Carr B., 1988, MNRAS 234, 219
Bahcall J.N., Hut P., Tremaine S., 1985, ApJ 290, 15
Bender R., Paquet A., Nieto J.-L., 1991, A&Ap 246, 349
Bender R., Burstein D., Faber S., 1992, ApJ 399, 462
Bender R., Burstein D., Faber S., 1993, in Panchromatic View of Galaxies: Their Evolutionary Puzzle, eds. G. Hensler et al., (Gif sur Yvette: Editions Frontières), in press

Bertin G., Stiavelli M., 1984, A&Ap 137, 26
Bertin G., Stiavelli M., 1993, Rep. Prog. Phys. 56, 493
Binggeli B., Cameron L.M., 1991, A&Ap 252, 27
Binney J., 1982, MNRAS 200, 951
Binney J., 1988, in The Word of Galaxies, eds. H. Corwin, Jr., L. Bottinelli, (New York: Springer Verlag), p. 332
Bothun G., Mould J., 1988, ApJ 324, 123
Burkert A., 1993, MNRAS, in press
Busarello G., Longo G., Feoli A., 1992, A&Ap 262, 52
Capaccioli M., 1988, in The Word of Galaxies, eds. H. Corwin, Jr., L. Bottinelli, (New York: Springer Verlag), p. 208
Capaccioli M., Caon N., D'Onofrio M., 1993, in The ESO/EIPC Workshop on Structure, Dynamics, and Chemical Evolution of Early Type Galaxies, eds. J. Danziger et al., ESO CWP #45, 43
Carlberg R., Lake G., Norman C., 1986, ApJ 300, L1
Carter D., Sadler E., 1990, MNRAS 245, 12
Ciotti L., 1991, A&Ap 249, 99
Daly R., McLaughlin G., 1992, ApJ 390, 423
Davies R., Efstathiou G., Fall S.M., Illingworth G., Schechter P., 1983, ApJ 266, 41
de Carvalho R., Djorgovski S., 1989, ApJ 341, L37
de Carvalho R., Djorgovski S., 1992a, ApJ 389, L49
de Carvalho R., Djorgovski S., 1992b, in Cosmology and Large-Scale Structure in the Universe, ed. R. de Carvalho, ASPCS, 24, 135
Dekel A., Silk J., 1986, ApJ 303, 39
de Zeeuw T., Franx M., 1991, ARAA 29, 239
Djorgovski S., 1985, Ph.D. Thesis, Univ. of California, Berkeley
Djorgovski S., King I.R., 1986, ApJ 305, L61
Djorgovski S., 1987, in Structure and Dynamics of Elliptical Galaxies, IAU Symp. #127, ed. T. de Zeeuw, (Dordrecht: D. Reidel), p. 79
Djorgovski S., 1988, in Starbursts and Galaxy Evolution, eds. T.X. Thuan et al., (Gif sur Yvette: Editions Frontières), p. 549
Djorgovski S., de Carvalho R., Han M.-S., 1988, in Extragalactic Distance Scale, eds. S. van den Bergh C. Pritchet, ASPCS, 4, 329
Djorgovski S., de Carvalho R., 1990, in Windows on Galaxies, eds. G. Fabbiano et al., (Dordrecht: Kluwer), p. 9
Djorgovski S., 1992a, in Morphological and Physical Classification of Galaxies, eds. G. Longo et al., (Dordrecht: Kluwer), p. 337
Djorgovski S., 1992b, in Cosmology and Large-Scale Structure in the Universe, ed. R. de Carvalho, ASPCS, 24, 19
Djorgovski S., 1992c, in Cosmology and Large-Scale Structure in the Universe, ed. R. de Carvalho, ASPCS, 24, 73
Djorgovski S., Bendinelli O., Parmeggiani G., Zavatti F., 1992, in Morphological and Physical Classification of Galaxies, eds. G. Longo et al., (Dordrecht: Kluwer), p. 439
Djorgovski S., 1993, in The Globular Cluster – Galaxy Connection, eds. G. Smith J. Brodie, ASPCS, 48, 496
Djorgovski S., Santiago B.X., 1993, in The ESO/EIPC Workshop on Structure, Dynamics, and Chemical Evolution of Early-Type Galaxies, eds. J. Danziger et al., ESO publication No. 45, 59
Faber S., Dressler A., Davies R., Burstein D., Lynden-Bell D., Terlevich R., Wegner G., 1987, in Nearly Normal Galaxies, ed. S. Faber, (New York: Springer Verlag), p. 175

Franx M., Illingworth G., de Zeeuw T., 1991, ApJ 383, 112
Gurzadyan V.G., Savvidy G.K., 1986, A&Ap 160, 203
Held E., Mould J., de Zeeuw T., 1990, AJ 100, 415
Hénon M., 1961, Ann. d'Aph. 24, 369
Kormendy J., 1982, in Morphology and Dynamics of Galaxies, 12th Saas-Fee Advanced Course, (Geneva: Geneva Observatory), p. 115
Kormendy J., 1985, ApJ 295, 73
Kormendy J., Djorgovski S., 1989, ARAA 27, 235
Lacey C., Ostriker J., 1985, ApJ 299, 633
Lynden-Bell D., 1967, MNRAS 136, 101
Maoz E., Bekenstein J., 1990, ApJ 353, 59
McGlynn T., 1984, ApJ 281, 13
Merritt D., 1988, AJ 95, 496
Mestel L., 1963, MNRAS 126, 553
Navarro J., 1993, in The ESO/EIPC Workshop on Structure, Dynamics, and Chemical Evolution of Early-Type Galaxies, eds. J. Danziger et al., ESO publication No. 45, 385
Peebles P.J.E., Dicke R.H., 1968, ApJ 154, 891
Penereiro J.C., Djorgovski S., de Carvalho R.R., Gorjian V., 1993, in The ESO/EIPC Workshop on Structure, Dynamics, and Chemical Evolution of Early-Type Galaxies, eds. J. Danziger et al., ESO publication No. 45, 487
Pfenniger D., 1986, A&Ap 165, 74
Pfenniger D., 1989, ApJ 343, 142
Renzini A., Ciotti L., 1993, ApJ, in press
Sandage A., Binggeli B., Tammann G., 1985, AJ 90, 1759
Schombert J., 1987, ApJS 64, 643
Silk J., Wyse R.F.G., 1993, Phys. Rep. 231, 293
Spergel D., Hernquist L., 1992, ApJ 397, L75
Stiavelli M., Bertin G., 1985, MNRAS 217, 735
Stiavelli M., Bertin G., 1987, MNRAS 229, 61
Tonry J., 1983, ApJ 266, 58
Tremaine S., Hénon M., Lynden-Bell D., 1986, MNRAS 219, 285
Tremaine S., 1976, ApJ 203, 345
Tremaine S., 1987, in Structure and Dynamics of Elliptical Galaxies, IAU Symp. #127, ed. T. de Zeeuw, (Dordrecht: D. Reidel), p. 367
van Albada T.S., 1982, MNRAS 201, 939
Vader J.P., 1986, ApJ 305, 669
Vader J.P., Vigroux L., Lachièze-Rey M., Souviron J., 1988, A&Ap 203, 217
Vigroux L., Souviron J., Vader J.P., 1984, A&Ap 139, L9
Villumsen J., 1984, ApJ 284, 75
White S., Narayan R., 1987, MNRAS 229, 103
Whitmore B., 1984, ApJ 278, 61
Wirth A., Gallagher J., 1984, ApJ 282, 85
Yoshii Y., Arimoto N., 1987, A&Ap 188, 13

Some Clues About the Dynamics of Globular Clusters from High-Resolution Observations

Georges MEYLAN

European Southern Observatory, Karl-Schwarzschild-Straße 2, D-8046 Garching bei München, Germany

Abstract. This review describes some of the high-resolution observations of the cores of globular clusters, obtained from the ground and with the Hubble Space Telescope (HST). Contrary to galaxies, whose distances are generally larger than 1 Mpc, implying unresolved stellar distributions, globular clusters have distances of the order of 10 kpc, which allow HST to resolve the brightest stars into grainy stellar distributions. We specifically discuss here some recent results on the galactic globular clusters NGC 6397, M15, and 47 Tucanae.

1. The Remaining Unique Capabilities of HST

It is now common knowledge that soon after launch, the in-orbit testing of the Hubble Space Telescope (HST) revealed a serious optical problem with its 2.4 m primary mirror, which suffers from a substantial amount of spherical aberration due to an error when it was ground more than a decade ago (Burrows et al. 1991). This gives rise to grossly defocused images in the telescope focal plane, at all settings of the secondary mirror.

The loss of energy in the core of the point spread function — the PSF is the image intensity distribution of a point source — is clearly visible in Fig. 1 (upper panel) where the simulated, originally expected, point spread function (left) is presented with the observed point spread function (right) obtained in orbit through the same filter, with the Faint Object Camera. Most of the light from the star is spread over a circle with a diameter of almost 5″. Fig. 1 (lower panel) shows that, apart from some energy in the core, not everything has been lost. The same simulated (left) and observed (right) point spread functions are displayed, but with a different upper intensity cut. It is conspicuous that a sharp core is still clearly present in the observed point spread function. It contains only 15% of the energy instead of the expected 80%, nevertheless, the diameter of this central core is only three pixels wide, which corresponds to 0.066″ (FWHM).

As a result of the flaw in the primary mirror, the point spread function of HST has a tight diffraction-limited core, as expected originally, surrounded

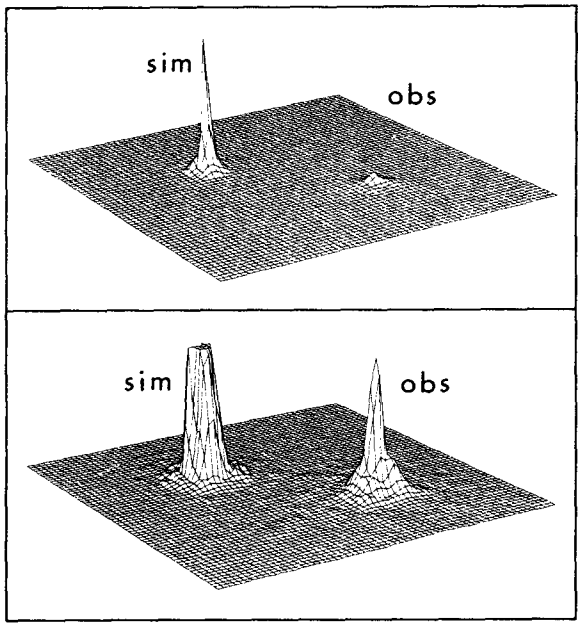


Fig. 1. The current problem and expectation of the Hubble Space Telescope (HST) are visualized in this figure. The upper panel displays the simulated, originally expected, point spread function (left) with its corresponding observed point spread function (right) obtained in orbit through the same filter, as observed by the Faint Object Camera. The loss in sensitivity is conspicuous. The lower panel shows the same data, but with a different upper cut. A sharp core is still present in the observed point spread function

by an unexpected extended halo due to the spherical aberration. Contrary to many statements, it is not so much the high spatial resolution of the Hubble Space Telescope cameras that has been compromised by the error in the mirror, but rather the limiting sensitivity: *HST has lost three magnitudes in sensitivity, but for bright enough objects, its* $\sim 0.1''$ *spatial resolving power is still present.* In addition, HST has retained its ultraviolet observational capabilities. This makes HST a useful telescope for the study of globular cluster cores.

2. NGC 6397: to Be or Not to Be Collapsed

Charge Coupled Device (CCD) observations have allowed, in the eighties, a systematic investigation of the inner surface brightness profiles of 127 Galactic globular clusters (Djorgovski & King 1986; Chernoff & Djorgovski 1989). These authors sorted the globular clusters into two different classes: *i*) the "King clusters", whose surface brightness profiles resemble a King model with a flat isothermal core and a steep envelope, and *ii*) the "collapsed-core clusters", whose surface brightness profiles follow an almost pure power law with an exponent of about −1. In the Galaxy, about 20% of the globular clusters belong to the second type, exhibiting in their inner regions apparent departures from King-model profiles. They are consequently considered to have collapsed cores.

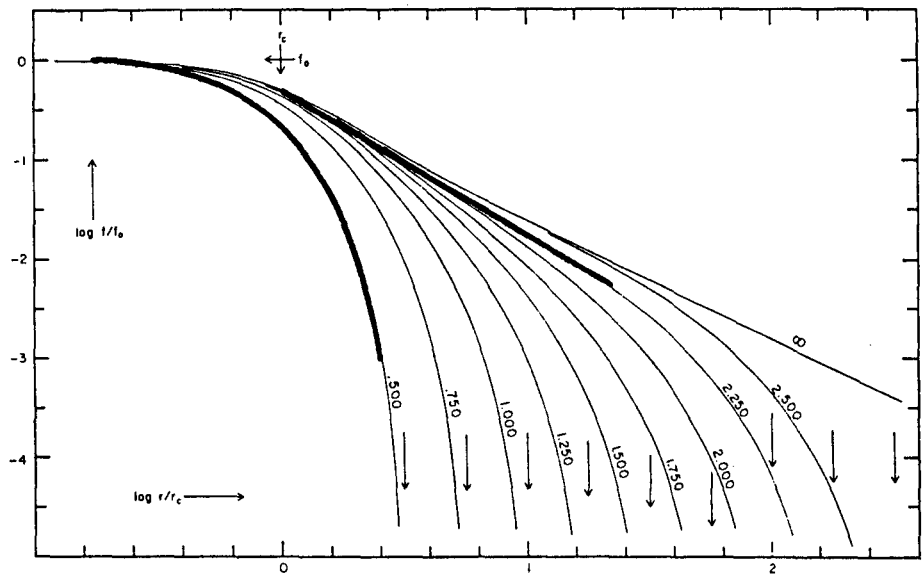


Fig. 2. Surface densities of the single-mass isotropic spherical models published by King (1966). Arrows indicate log r_t for models with concentrations $c = \log(r_t/r_c)$ in the range $0.5 - 2.5$. The observable part of a low-concentration model –"King clusters"– corresponds to a flat isothermal core and a steep envelope, when the observable part of a high-concentration model –"collapsed-core clusters"– corresponds to an almost pure power law with an exponent of about -1

The distinction between King and collapsed-core clusters is not always clear (see, e.g., Meylan & Pryor 1993, and references therein). Integrated surface brightnesses measured for small areas in the cores of globular clusters are strongly dominated by statistical fluctuations in the small numbers of bright stars within the aperture. Does the surface brightness profile of NGC 6397 show any sign of core collapse? Maybe, since it increases toward the center and is bumpy in its inner 100″. This is especially true when observed through B or U filters, because of a high concentration of blue stragglers (Aurière et al. 1990), which contribute strongly to the color gradient. However, CCD observations in the core of NGC 6397, aperture photometry at intermediate radii, and star counts at large radii allow the construction of a surface brightness profile extending from the core out to 25′. The multi-mass King-Michie model which fits this surface brightness profile reasonably well has a very high concentration, viz., $c = \log(r_t/r_c) \simeq 2.5$, where r_t and r_c are the tidal and core radii, respectively (Meylan & Mayor 1991). Such a large value of the concentration is a clear indication that NGC 6397 is dynamically highly evolved.

In a similar way, Grabhorn et al. (1992) are able to fit successfully a multi-mass King model of even higher concentration, viz., $c \simeq 3.0$, to the surface

brightness profile of M15, the prototype of the collapsed-core globular clusters. These two successful fits may indicate that the dynamical status of a cluster should be discussed in the context of the whole surface brightness profile, and may be deduced from the value of its concentration. The power-law shape of the inner observable part of a high-concentration profile, a power-law predicted by King and Fokker-Planck models may be difficult to observe being subject to statistical fluctuations in the small numbers of bright stars. Consequently, any globular cluster with a concentration $c = \log(r_t/r_c) \gtrsim 2.0 - 2.5$ may be considered as collapsed, or on the verge of collapsing, or just beyond. (It is worth mentioning that the pre-, in-, and post-collapse terminology has only a theoretical meaning, since observations are unable to differentiate these three phases). The apparent discrepancy between "King or power-law" and "concentration" criteria may be reconciled with a look at Fig. 2, which displays, for different concentrations $c = \log(r_t/r_c)$, a grid of models (King 1966) which incorporate the two most important elements governing globular cluster structure: two-body relaxation and tidal truncation. We see that high-concentration models have profiles which, in their inner observable part, are characterized by power-laws.

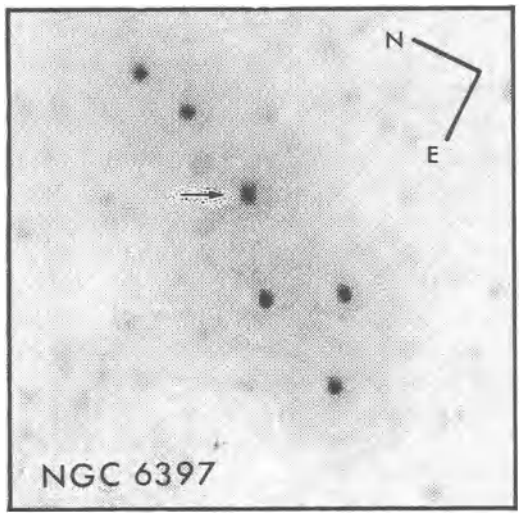


Fig. 3. Part of an image of the core of NGC 6397 taken with the F/48 mode, through the F220W filter (2250 Å) of the Faint Object Camera on board the Hubble Space Telescope. The size of the image is 11″ × 11″ (Burgarella et al. 1993). The luminosity center of the cluster is within 1″ from the star marked by the arrow

The recent HST observations of the core of NGC 6397, with the Faint Object Camera, fully resolve the core of this globular cluster and rule out definitely the presence of a bright central cusp in luminosity, a possibility not entirely eliminated by ground-based observations. The inner few seconds of arc, displayed in Fig. 3, are dominated by the presence of 6 bright blue stragglers, which, for a globular cluster, constitute a prominent concentration of such stars (Burgarella et al. 1993).

Given its very high concentration, and independently of the shape of its inner surface brightness profile, NGC 6397 may be considered as core collapsed. The presence of the blue stragglers, possibly formed by mergers after collisions between two stars, may very well be a direct consequence of core collapse and its associated very high stellar density. Some indications of the fact that these stars may have individual stellar masses higher than the turn-off mass (i.e., $\gtrsim 0.8\,M_\odot$) are found in the more centrally-concentrated distribution of the blue stragglers with respect to the subgiants of same magnitudes, possibly pointing towards the presence of mass segregation (Lauzeral et al. 1992; see also Burgarella et al. 1993). There is no observed dynamical signature of core collapse from velocity dispersion measurements (Meylan & Mayor 1991; Dubath, Meylan, & Mayor 1993a), a dynamical signature which would be difficult to observe, anyway, according to theoretical predictions by, e.g., Grabhorn et al. (1992).

3. M15: a Prototype Collapsed-Core Cluster?

The globular cluster M15 = NGC 7078 has long been considered as a prototype of the collapsed-core star clusters. Early electronographic determinations of its luminosity profile by Newell & O'Neil (1978), confirmed by further photographic and CCD studies (e.g., Aurière & Cordoni 1981), reveal a central excess of light. Newell, Da Costa, & Norris (1976) found that these observations were consistent with the existence of a central massive object, possibly a black hole of about $800\,M_\odot$, while Illingworth & King (1977) were able to successfully fit dynamical models to the entire surface-brightness profile without invoking a black hole. The latter authors explained the central brightness peak as being caused by the gravitational effect of a centrally-concentrated population of neutron stars.

Cudworth's (1976) proper motion study gave the first estimate of velocity dispersion in M15, $\sigma_p \sim 10\,km\,s^{-1}$, based on stars between 1.5′ and 12′ from the center. The only other published work providing comprehensive observational constraints on the dynamics of M15 is Peterson, Seitzer, & Cudworth (1989). In their study, the velocity dispersion is derived from two different kinds of data: *i*) from individual radial velocities for 120 cluster members scattered between 0.1′ and 4.6′ from the center and *ii*) from integrated-light spectra of the central luminosity cusp. The radial velocities of the 27 stars within 20″ of the center give $\sigma_p = 14.2 \pm 1.9\,km\,s^{-1}$, while the integrated-light spectra suggest a cusp in the velocity dispersion profile, with $\sigma_p(0)$ of at least $25\,km\,s^{-1}$. According to the work of Peterson et al., the core of M15 exhibits a sharp rise in velocity dispersion within the central few seconds of arc, a unique case among globular clusters. Peterson et al. (1989) find that this cusp in the velocity dispersion is consistent with neither the King-Michie models nor the post-core collapse models. They suggest that their observations indicate a nonthermal velocity distribution, consistent with a central

black hole of about $1,000\,M_\odot$. It is worth mentioning that, from the radial distribution of the pulsars in M15, Phinney (1993) deduces that this cluster does not contain more than $\sim 10^2$ stellar mass black holes.

Since the study of Peterson et al. (1989), high-resolution imaging of the center of M15 has resolved the luminosity cusp into stars (the cusp was already partly resolved in the images published by Aurière et al. 1984). Images with a FWHM resolution of 0.35″, taken by Racine & McClure (1989) with the High-Resolution Camera of the Canada-France-Hawaii Telescope (CFHT), and, in particular, images with a FWHM of 0.08″ obtained with the Planetary Camera of the Hubble Space Telescope (HST) by Lauer et al. (1991) show that the cusp is easily resolved into a group of a few bright stars. On one hand, Lauer et al. (1991) show that the surface-brightness profile of the residual light, obtained after subtracting the bright resolved stars, does not continue to rise at subarcsecond radii, but flattens off interior to a radius of about 2″. They determine a core radius of $2.2'' = 0.13\,\mathrm{pc}$ from their observations and argue that the existence of a core, rather than a cusp, at the center of M15 indicates that the central dark matter implied by the high central velocity dispersion measured by Peterson et al. (1989) does not belong to a massive black hole, but probably resides in a more diffuse form. On the other hand, from their own HST data, Yanny et al. (1993) find that a flat core is not apparent for $r \gtrsim 1.5''$. They find the radial distribution consistent with a number of scenarios, including: *i*) a central black hole of mass $> 1000\,M_\odot$; *ii*) a collapsed core with steep central profile $\alpha < -0.75$, and *iii*) a small flat core of radius $\lesssim 1.5'' = 0.09\,\mathrm{pc}$. Earlier reports of weak color gradient in the center of M15 (Bailyn et al. 1988) are confirmed in the sense that bright red giants are depleted in the center relative to subgiants, but the depletion of very blue HB stars counteracts this bluing (Stetson 1991).

Recently, but see also King (1989), Grabhorn et al. (1992) successfully fit the high-resolution surface-brightness profile determined by Lauer et al. (1991) with the profile predicted by multi-mass King and Fokker-Planck models. Their predicted velocity dispersion profile from Fokker-Planck models matches the observations of Peterson et al. (1989) reasonably well, except for the $\sigma_p(0) = 25\,\mathrm{km\,s^{-1}}$ central value. This is conspicuous in Fig. 4, from Grabhorn et al. (1992), where the velocity dispersion profile, even in deep core collapse, never reaches the Peterson et al.'s high value. A simple extrapolation of the deep core collapse profile (Fig. 4) reaches $\sigma_p(0) = 25\,\mathrm{km\,s^{-1}}$ only within 0.006″, i.e., over an area significantly smaller than the 1″ sampling area of Peterson et al. (1989). Do we really need a massive black hole to explain the observations of M15 ? Are the results of Peterson et al. (1989) statistically significant ?

As part of a long-term program to determine the central velocity dispersion in the cores of high-concentration and collapsed-core globular clusters, Dubath, Meylan, & Mayor (1993a,b) obtained, at the European Southern Observatory (ESO) at La Silla, Chile, an integrated-light spectrum of the

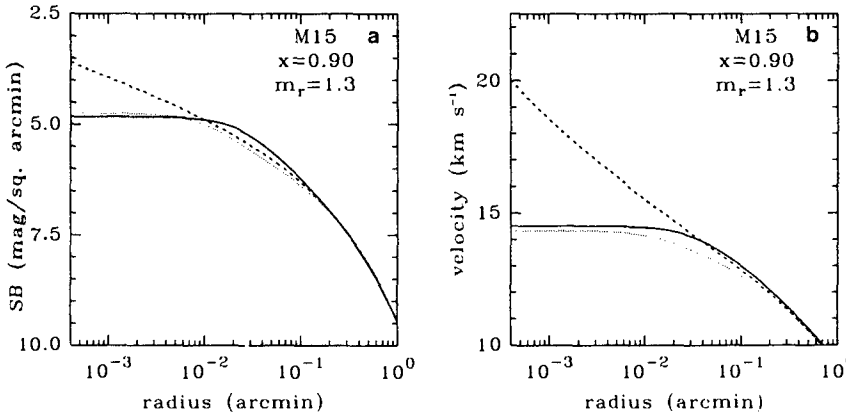


Fig. 4. Surface-brightness and velocity dispersion profiles at different times during postcollapse evolution. Solid curve: precollapse; dashed curve: core-collapse; dotted curve: postcollapse (from Grabhorn et al. 1992)

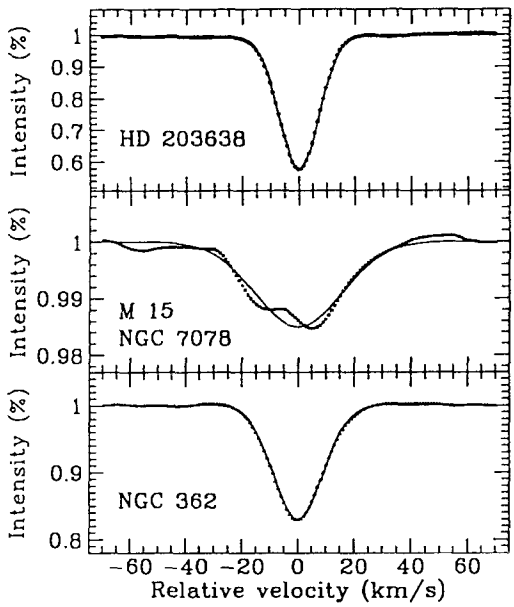


Fig. 5. Normalized cross-correlation functions, *i*) of the spectrum of HD 203638, a standard K0 giant star, *ii*) of the integrated light spectrum of the central $6'' \times 6''$ area in the core of M15 = NGC 7078, and *iii*) of the integrated light spectrum of the central $6'' \times 6''$ area in the core of another galactic globular cluster, NGC 362. The continuous lines are the corresponding fitted Gaussians. The important broadening of both cluster cross-correlation functions as well as the asymmetrical shape of the M15 cross-correlation function are conspicuous (Dubath et al. 1993b)

core of M15. Fig. 5 displays the cross-correlation function (CCF) of the spectrum of a standard K0 giant star, of the integrated-light spectrum of the core of M15, and of the CCF of the integrated-light spectrum of the core of the Galactic globular cluster NGC 362, all three spectra having been obtained under identical conditions. The CCF's of the spectra of HD 203638 and of

NGC 362 are displayed in order to show how similar to a Gaussian the typical CCF's obtained for standard stars and for other Galactic globular clusters can be. Totally unexpectedly, and despite the high signal-to-noise ratio of the observed spectrum, the cross-correlation function of the M15 spectrum is bumpy, as if it were the sum of two different Gaussians. This large departure from the usual Gaussian function is larger than the deviations produced by the spectrum noise. Such a behavior (also present in the CCF of Peterson et al. 1989) of the cross-correlation function is expected only if the integrated-light spectrum is completely dominated by the contribution of the few brightest stars lying inside the sampling area (slit) of the spectrograph.

The significant Doppler broadening, due to the spatial random motions of the stars, of the CCF's of both clusters is conspicuous when compared with the stellar CCF of HD 203638. The standard deviation of the intrinsic stellar CCF ($\sigma_{\rm ref} = 7.0\,{\rm km\,s^{-1}}$) is defined by the mean value of the measurements of a sample of standard stars of appropriate spectral type. The velocity dispersion $\sigma_{\rm p}$ is then computed with ${\sigma_{\rm p}}^2 = {\sigma_{\rm CCF}}^2 - {\sigma_{\rm ref}}^2$. The integrated-light spectrum of Dubath, Meylan, & Mayor (1993b) over a central $6'' \times 6''$ area leads to a projected velocity dispersion $\sigma_{\rm p}(0) = 14.0\,{\rm km\,s^{-1}}$.

It is worth mentioning that, because of a larger sampling area, Dubath et al. would probably miss any central cusp in velocity dispersion. The important point here comes from the bumpy shape of both Dubath et al.'s and Peterson et al.'s CCFs, pointing towards probable sampling problems. A quantitative estimate of the sampling uncertainties affecting the central velocity dispersion measurements of M15 is absolutely necessary for further interpretations of any results. Detailed and exhaustive numerical simulations, with different sampling apertures ($1'' \times 1''$ in the case of Peterson et al., and $6'' \times 6''$ in the case of Dubath et al.), of the CCF's of integrated-light spectra in the core of M15 have been carried out by Dubath et al. (1993b). The results may be summarized by two points: *i*) The noisy shapes of the Peterson et al.'s and Dubath et al.'s observed CCFs of M15 are qualitatively easily reproduced by the simulations. *ii*) with the above $\sigma_{\rm p} = 14\,{\rm km\,s^{-1}}$ as input data, the value $\sigma_{\rm p} = 15\,^{+6}_{-4}\,{\rm km\,s^{-1}}$ is obtained over an area of integration of $6'' \times 6''$.

A precise estimate of the sampling errors in the results of Peterson et al. (1989) cannot be directly deduced from Dubath et al.'s simulations, the two cross-correlation techniques being not exactly the same. Nevertheless, Dubath et al.'s simulations adapted to Peterson et al.'s case suggest that the sampling errors in their results are much larger than in Dubath et al.'s case and larger than what has been acknowledged so far (see also Zaggia et al. 1993 and references therein). The observational evidence for a central cusp in velocity dispersion in the core of M15 is therefore not convincing. There are also some inconsistencies in the radial velocities obtained by Peterson et al. (1989) from their different velocity dispersion values. Dubath et al.'s simulations indicate that a very accurate measurement of the velocity dispersion within the central few seconds of arc in the core of M15 cannot be achieved from integrated-

light measurement. Further work is clearly needed in order to confirm, or disprove, the significance of the large central velocity dispersion measured by Peterson et al. (1989) in the core of M15. Observations with a Fabry-Pérot interferometer will be of great interest, and will perhaps be able to solve this enigma. It would be very interesting to get some observational evidences that M15 may very well be one of the few galactic globular clusters caught in a state of deep core collapse.

4. The Rich Stellar Phenomenology of 47 TUC

For decades, 47 Tucanae has been known as one of the most massive globular clusters in our Galaxy. Well observed and well studied, photometrically and kinematically, it has a rather high concentration, viz., $c \simeq 2.1$, but is perfectly well fitted by multi-mass anisotropic King-Michie models and may be considered as slowly evolving through quasi-equilibrium states towards its ultimate fate of core collapse (Meylan 1989). During the last few years, there has been an increasing amount of direct and indirect observational evidence for stellar encounters and collisions in the core of 47 Tucanae, including the presence of:

1. eleven millisecond pulsars, some of them binaries (Manchester et al. 1990 and 1991),
2. blue stragglers within 20″ of the core (Paresce et al. 1991),
3. two high-velocity stars ejected out of the core (Meylan, Dubath, & Mayor 1991),
4. four X-ray sources detected by ROSAT (Verbunt et al. 1993).

Similar evidence of ongoing stellar phenomenology are also found in other globular clusters, though in more limited numbers. Another way to estimate the effect of dynamical and stellar evolutions comes from color and population gradient studies (Djorgovski & Piotto 1993). The core of 47 Tucanae, with $r_c \simeq 25''$ (Meylan 1989), was resolved a long time ago from the ground. Two recent studies, using HST observations have investigated, at high spatial resolution, the inner structure of this cluster, looking for color and population gradients and for possible signs of core collapse or the presence of a central massive object.

Calzetti et al. (1993), using observations taken with the HST Faint Object Camera, have investigated the position of the center of the cluster. They determine, first, the "center of luminosity" with the mirror autocorrelation technique, and, second, the "center of gravity" with the mirror autocorrelation technique applied not to the whole light but to the stars considered point-like with equal weights. The difference of 6″ between the two center positions would be an interesting observation, but it is actually of the order of the uncertainty on the position of the center of luminosity as determined by autocorrelation technique. Assuming that their center of gravity is the

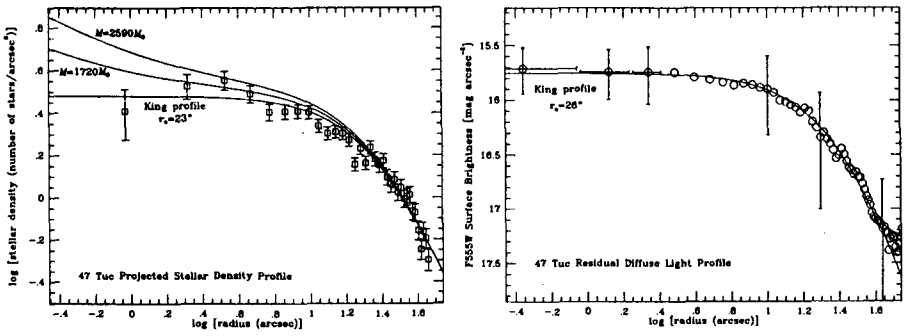


Fig. 6. Left: The projected density profile of all stars observed in 47 Tucanae by Guhathakurta et al. (1992). The best fitting King profile is shown by the lowest of the three curves. The two upper curves, evolving central compact objects, are ruled out at the 90% and 95% confidence levels.
Right: The radial surface brightness profile of the residual light well fitted by a King profile. Following Guhathakurta et al. (1992), there is no evidence for luminosity cusp in 47 Tucanae

true center of the cluster, these authors find a density profile which shows a central cusp with a core radius of about 8″ and which deviates significantly from a King profile up to 13″ from the center.

The above results are in disagreement with those from Guhathakurta et al. (1992). Using observations taken with the HST Planetary Camera, they show that the region within 1′ from the center contains a centrally concentrated population of blue stragglers. The radial profile of the projected stellar density (giants) is flat in the central region, with a core radius $r_c = 23'' \pm 2''$ (Fig. 6 left). No signature of a collapsed core is evident, as already noticed by Djorgovski & King (1984). The surface brightness distribution of the diffuse residual light has a core radius $r_c = 26''$ (Fig. 6 right) that is marginally larger than that of the giant stars. The difference between these two core radius values may be an indication of mass segregation, since the faint stars at or just below the main sequence turn-off that contribute most of the residual light are slightly less massive than the giants.

Among the galactic globular clusters studied in detail so far, 47 Tucanae contains by far the largest amount of indirect evidence of stellar collisions and encounters. There is no dynamical signature of core collapse, either from surface brightness profile or from velocity dispersion profile, a signature not expected anyway given the concentration of 47 Tucanae only marginally high. The two high-velocity stars ejected out of the core may be indications that 47 Tucanae is burning its primordial binaries in order to delay its core collapse.

5. Conclusions

At the present stage of understanding of globular cluster dynamical evolution, we need more sophisticated models constrained by more numerous data of better quality, as, e.g., high-quality star counts, proper motions, and radial velocities. During the last few years, numerous studies of high-concentration globular clusters, with $c = \log(r_t/r_c) \gtrsim 2$, have confirmed observationally what was strongly suspected: stellar evolution and stellar dynamics are intimately connected. Studies concerning individual stars as well as those devoted to integrated properties of stellar distributions (e.g., color and population gradients) show that stellar encounters, collisions, mergers, complicate and enrich the dynamical study of globular clusters, considered as simple systems not long ago!

Acknowledgements. It is a pleasure to thank my collaborators, P. Dubath and M. Mayor, for allowing some of our data to be quoted before publication, and S.G. Djorgovski for some interesting discussions and comments.

References

Aurière M., Cordoni J.-P., 1981, Astr. Ap. 100, 307
Aurière M., Le Fèvre O., Terzan A., 1984, Astr. Ap. 138, 415
Aurière M., Ortolani S., Lauzeral C., 1990, Nature 344, 638
Bailyn C.D., Grindlay J.E., Cohn H., Lugger P.M., 1988, Ap. J. 331, 303
Burgarella D., et al. 1993, Astr. Ap. submitted
Burrows C. J., Holtzman J. A., Faber S. M., Bely P. Y., Hasan H., Lynds C. R., Schroeder D., 1991, Ap. J. (Letters) 369, L 21
Calzetti D., De Marchi G., Paresce F., Shara M., 1993, Ap. J. (Letters) 402, L 1
Chernoff D.F., Djorgovski S., 1989, Ap. J. 339, 904
Cudworth K.M., 1976, A. J. 81, 519
Djorgovski S., King I.R., 1984, Ap. J. (Letters) 277, L 49
Djorgovski S., King I.R., 1986, Ap. J. (Letters) 305, L 61
Djorgovski S.G., Piotto G., 1993 in ASP Conf. Series Vol. 50 on *Structure and Dynamics of Globular Clusters*, eds. S. Djorgovski, G. Meylan (San Francisco: Astronomical Society of the Pacific), in press
Dubath P., Meylan G., Mayor M., 1993a, Astr. Ap. submitted
Dubath P., Meylan G., Mayor M., 1993b, Ap. J. submitted
Grabhorn R.P., Cohn H.N., Lugger P.M., Murphy B.W., 1992, Ap. J. 392, 86
Guhathakurta P., Yanny B., Schneider D.P., Bahcall J.N. 1992, A. J. 104, 1790
Illingworth G., King I.R., 1977, Ap. J. 218, L109
King I.R., 1966, A. J. 71, 64
King I.R., 1989, in *Dynamics of Dense Stellar Systems*, ed. D. Merritt (Cambridge: Cambridge University Press), p. 157
Lauer T.R., et al. 1991, Ap. J. 369, L45
Lauzeral C., Ortolani S., Aurière M., Melnick J., 1992, Astr. Ap. 262, 63
Manchester R.N., et al. 1990, Nature 345, 598
Manchester R.N., et al. 1991, Nature 352, 219
Meylan G., 1989, Astr. Ap. 214, 106
Meylan G., Mayor M., 1991, Astr. Ap. 250, 113

Meylan G., Dubath P., Mayor M., 1991, Ap. J. 383, 587
Meylan G., Pryor C., 1993 in ASP Conf. Series Vol. 50 on *Structure and Dynamics of Globular Clusters*, eds. S. Djorgovski, G. Meylan (San Francisco: Astronomical Society of the Pacific), in press
Newell B., Da Costa G.S., Norris J., 1976, Ap. J. 208, L55
Newell B., O'Neil E.J., 1978, Ap. J. Suppl. 37, 27
Paresce F., et al. 1991, Nature 352, 297
Peterson R.C., Seitzer P., Cudworth K.M., 1989, Ap. J. 347 251
Phinney E.S., 1993 in ASP Conf. Series Vol. 50 on *Structure and Dynamics of Globular Clusters*, eds. S. Djorgovski, G. Meylan (San Francisco: Astronomical Society of the Pacific), in press
Racine R., McClure R.D., 1989, Pub. A.S.P. 101, 731
Stetson P., 1991 in *Precision Photometry: Astrophysics of the Galaxy*, eds. A.G.D. Philip, A.R. Upgren, K. Janes (Schenectady: L. Davis Press), p. 69
Verbunt F., Hasinger G., Johnston H.M., Bunk W., 1993 in *Recent Results in X-ray and EUV Astronomy*, eds. (: Press), in press; (also Preprint Sterrekundig Instituut Utrecht #920008)
Yanny B., Guhathakurta P., Schneider D.P., Bahcall J.N. 1993, Ap. J. (Letters) in press
Zaggia S.R., Capaccioli M., Piotto G., 1993 in ASP Conf. Series Vol. 50 on *Structure and Dynamics of Globular Clusters*, eds. S. Djorgovski, G. Meylan (San Francisco: Astronomical Society of the Pacific), in press

Diffusion of Stellar Orbits in the Galactic Disk

B. FUCHS, C. DETTBARN, R. WIELEN

Astronomisches Rechen-Institut Heidelberg, Germany

1. Observational Evidence

Direct observational evidence for a diffusion of stellar orbits in velocity space is provided by observations of the velocity dispersions of stars in the solar neighbourhood (Wielen 1977; Wielen et al. 1992) which show a marked increase with stellar ages. This could be either due to higher velocity dispersions of the interstellar gas, out of which the stars were born, at earlier epochs of galactic evolution or an increase of the velocity dispersions of the stars during their lifetime. We have argued before that if the increase of the velocity dispersion solely reflects the velocity dispersions with which the stars were born, this would mean an undue preference of the present epoch of galactic evolution, since the velocity dispersions of the stars would have then dropped very rapidly over the last $2 \cdot 10^9$ years (cf. Fig. 1a). We conclude thus, that the stars have experienced stochastic accelerations by inhomogeneities of the galactic gravitational field, leading to a continuous increase of the velocity dispersions with stellar ages.

2. Physical Mechanisms

Any inhomogeneities in the matter distribution in the galactic disk imply statistical fluctuations of the galactic gravitational force field, g_μ, which perturb the stellar orbits in the galactic disk and will eventually lead to an increase of the stellar velocity dispersions. The increase may be described quantitatively by diffusion coefficients and for weak fluctuations these are quite generally given by (Fuchs 1980)

$$D_{\mu\lambda} = \int_{-\infty}^{t} \mathrm{d}t' \sum\nolimits_\nu \left\langle g_\mu(\mathbf{X}(t),t)\, g_\nu(\mathbf{X}(t'),t') \right\rangle \frac{\partial X_\lambda(t)}{\partial X_\nu(t')} \,, \tag{1}$$

where $\mathbf{X}(t)$ denotes the phase space coordinates of the unperturbed stellar orbits. In the case of gravitational encounters of the kind we consider here there are no spatial elements of $D_{\mu\lambda}$.

However, in order to make specific predictions for the evolution of the velocity distribution of the stars under the influence of the perturbers a number of concrete physical mechanisms of stochastic acceleration of stars in the galactic disk have been discussed in the literature. Among these are

(a) deflections of stellar orbits by giant molecular clouds (Spitzer & Schwarzschild 1951, 1953; Lacey 1984),
(b) scattering of stars by massive black holes from the galactic halo (Lacey & Ostriker 1985),
(c) stochastic heating of the galactic disk by transient spiral density wavelets (Sellwood & Carlberg 1984; Jenkins & Binney 1990), or
(d) heating of the galactic disk by infalling, disrupting satellite galaxies (Tóth & Ostriker 1992).

All these mechanisms are probably responsible to some degree for the diffusion of stellar orbits in the galactic disk. But the physical nature of the diffusion driven by these mechanisms is at present only well understood in the first two cases so that we concentrate here on these.

Spitzer & Schwarzschild (1951, 1953) assume plane galactic orbits in epicyclic approximation. The gravitational encounters between the stars and the giant molecular clouds lead to accelerations of the stars which are mere random deflections of the stellar orbits but leave the speed of the stars unchanged. The size of an epicycle is determined by the radial action integral or "epicyclic energy", $J_r = U^2 + (\kappa^2/4B^2)V^2$, with U the radial and V the azimuthal velocity components, while κ and B denote the epicyclic frequency and Oort's constant, respectively. If the encounters are assumed to be instantaneous the average gain of radial action per unit time is (Wielen & Fuchs 1983)

$$\langle \delta J_r \rangle = \left(\frac{\kappa^2}{4B^2} - 1 \right) (U^2 - V^2) \left\langle (\delta\Theta)^2 \right\rangle \tag{2}$$

with the 3-dimensional mean square deflection $\left\langle (\delta\Theta)^2 \right\rangle = 4\pi G^2 n_c m_c^2 \ln \alpha \Delta\tau / (U^2 + V^2)^{3/2}$, where n_c, m_c are the spatial density and the mass of the clouds, respectively (Spitzer & Schwarzschild 1953). Averaged over one epicyclic period the average gain of radial action is

$$\frac{\mathrm{d}J_r}{\mathrm{d}\tau} \propto \left(\frac{\kappa^2}{4B^2} - 1 \right) \frac{1}{\sqrt{J_r}} \quad \text{so that} \quad J_r \propto \tau^{2/3} \tag{3}$$

and the velocity dispersions grow as $\langle U^2 \rangle \propto \langle V^2 \rangle \propto J_r$. Note that the mechanism works only for non-circular epicycles, $\langle U^2 \rangle \neq \langle V^2 \rangle$ or $\kappa^2/4B^2 \neq 1$. The mechanism is not very efficient in accelerating stars. Assuming $m_c = 5 \cdot 10^5 \mathrm{M}_\odot$ and $n_c m_c = 0.02\,\mathrm{M}_\odot/\mathrm{pc}^3$, as in the solar neighbourhood, and a flat

galactic rotation curve, $\kappa^2/4B^2 = 2$, the radial velocity dispersion of stars of the age of 10^{10} yr would be only 29 km/s if this were the only diffusion process, whereas the observed radial velocity dispersion is about 55 km/s. However, there is a very considerable spatial dispersion related to this process. The rms deviation from the mean radius increases as $\langle (r - r_0)^2 \rangle = 26.2 \langle U^2 \rangle / \kappa^2$, assuming again a flat rotation curve. At a mean distance of $r_0 = 8.5$ kpc from the galactic center this implies for stars with an age of 10^{10} yr a rms radial deviation of 4 kpc from their mean radius.

In contrast to the giant molecular clouds with their small peculiar velocities massive black holes from the galactic halo would have such high velocities when they penetrate the galactic disk that they could actually change the speed of stars by gravitational encounters (Lacey & Ostriker 1985). The random accelerations add up statistically which may be described again quantitatively by diffusion coefficients, $\sum_i \langle (\Delta U_i)^2 \rangle = D_U \, \mathrm{d}\tau$, $\sum_i \langle (\Delta V_i)^2 \rangle = D_V \, \mathrm{d}\tau$, $\sum_i \langle (\Delta W_i)^2 \rangle = D_W \, \mathrm{d}\tau$, where W denotes the vertical stellar velocity component. In this case we have

$$\frac{\mathrm{d}J_r}{\mathrm{d}\tau} = D_U + \frac{\kappa^2}{4B^2} D_V \quad \text{so that} \quad J_r \propto \tau \tag{4}$$

and again the velocity dispersions grow as $\langle U^2 \rangle \propto \langle V^2 \rangle \propto \langle W^2 \rangle \propto J_r$. Wielen (1977) has shown that such relations may be fit nearly ideally to the observations allowing thus the empirical determination of the diffusion coefficients as $D_U = D_V = D_W = 2 \cdot 10^{-7} (\mathrm{km/s})^2/\mathrm{yr}$. The spatial diffusion related to this process is not as pronounced as in the Spitzer-Schwarzschild mechanism. The rms radial deviation from the mean galactocentric radius increases only as $\langle (r - r_0)^2 \rangle = 2.3 \langle U^2 \rangle / \kappa^2$ in the case of a flat rotation curve.

3. Numerical Simulations

The discussion of the diffusion of stellar orbits as outlined above involves a number of approximations, the most restrictive being the approximation of instantaneous acceleration by gravitational encounters. In order to avoid such approximations we have performed a series of numerical simulations of the diffusion of stars in the galactic disk. Orbits of stars are calculated by numerical integration in a rotating frame representing a patch of the galactic disk which is populated by various perturbers modelled as small Plummer spheres. The distribution of these objects is extended periodically in neighbouring frames which slide along the central frame according to the differential galactic rotation. We are thus able to follow the time evolution of the three-dimensional velocity distribution function of the stars as well as the evolution of the vertical structure of the disk.

Results for the 3-dimensional velocity dispersions are shown in Fig. 1a in comparison with the observed velocity dispersions of stars in the solar

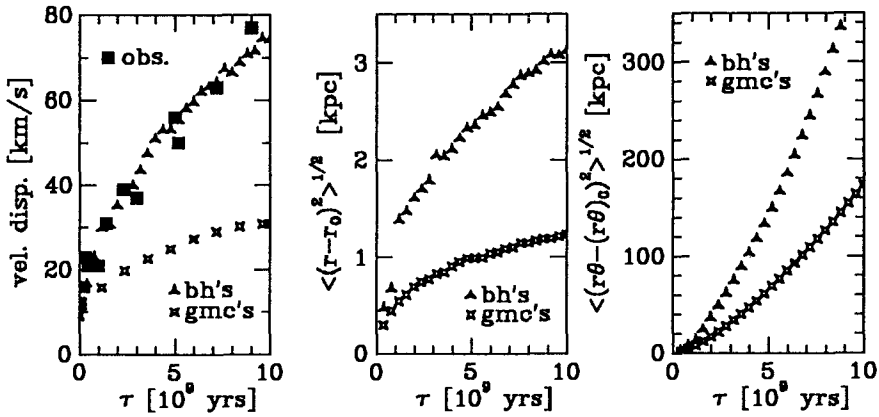


Fig. 1. Velocity dispersion (**a**, left), radial diffusion (**b**, middle), and tangential diffusion (**c**, right) of black-holes and giant molecular clouds as a function of time

neighbourhood. The black holes are assumed to have masses of $5 \cdot 10^6\,M_\odot$ and to be distributed homogeneously with a density of $0.01\,M_\odot/pc^3$. Their orbits are randomly oriented straight orbits with a typical velocity of 155 km/s. Their Plummer radii are set arbitrarily to 10^{-6} pc. The giant molecular clouds with Plummer radii of 20 pc, on the other hand, are assumed to have masses of $5 \cdot 10^5\,M_\odot$ and to be distributed in a narrow disk with a vertical scale height of 70 pc and a density of $0.02\,M_\odot/pc^3$ at the midplane. They orbit on epicyclic orbits corresponding to a 3-dimensional velocity dispersion of 5 km/s. As expected the diffusion driven by black holes leads to a velocity-dispersion-age relation which fits well to the observations, but the diffusion driven by giant molecular clouds is much too inefficient. Due to the small scale height of the molecular cloud distribution the slope of velocity-dispersion-age relation is even flatter than a $\tau^{1/3}$-law (cf. Lacey 1984) and does not fit to the observations even if the masses of the molecular clouds were increased. In Figs. 1b-c the corresponding spatial diffusion is illustrated. Both processes lead to a marked radial dispersion, but in particular to a dispersion in the tangential direction. This is due to the scattering of stars from their original epicycles onto epicycles with different guiding center motions so that they drift rapidly away in the tangential direction.

The mean galactocentric radius r_0 of a stellar orbit, however, is left unchanged by the diffusion processes. This implies that the exponential radial density profile of the galactic disk is not changed by the diffusion of stellar orbits. But the spatial dispersion will lead to an enhanced mixing of stars from different parts of the galactic disk. This effect is clearly seen in the dispersion of the metallicities of stars of a given age and at a certain galactocentric radius. Assuming that the mean radial variation of the metallicity of the stars

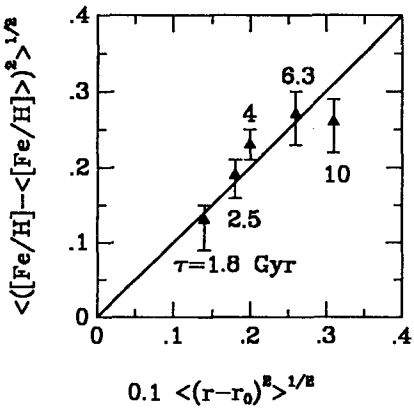


Fig. 2. Metallicity dispersion of various groups of stars of different ages as a function of the spatial dispersion

may be approximated as $\langle [\mathrm{Fe/H}] \rangle_r = \langle [\mathrm{Fe/H}] \rangle_{r_0} + \frac{\mathrm{d}}{\mathrm{d}r} \langle [\mathrm{Fe/H}] \rangle \cdot (r - r_0)$, the dispersion of the metallicities is related to the spatial dispersion as

$$\left\langle \left([\mathrm{Fe/H}] - \langle [\mathrm{Fe/H}] \rangle \right)^2 \right\rangle^{1/2} = \frac{\mathrm{d}}{\mathrm{d}r} \langle [\mathrm{Fe/H}] \rangle \cdot \left\langle (r - r_0)^2 \right\rangle^{1/2}. \quad (5)$$

In Fig. 2 metallicity dispersions of various groups of stars of different ages are shown. The metallicity dispersions which we predict using the spatial dispersion of stellar orbits found for the diffusion process driven by massive black holes (cf. Fig. 1b) and adopting a mean gradient of 0.1 kpc^{-1} (Grenon 1987) are drawn versus metallicity dispersions observed in the solar neighbourhood (Edvardson et al. 1993, cited in Freeman 1991). The error bars indicate statistical uncertainties. The individual accuracy of the observed metallicities is about 0.1 dex. The close correlation between the predicted and the actually observed metallicity dispersions indicates that the spatial dispersion of stellar orbits due to the diffusion process contributes significantly to the scatter of the metallicities of the stars in the solar neighbourhood.

4. Discussion

The scenario of massive black holes from the galactic halo driving the diffusion of stellar orbits explains successfully many of the observed characteristics of the distribution of the disk stars in phase space, not just only the age-velocity dispersion-relation but also the *shape* of the velocity distribution or the global variation of the velocity dispersion as function of galactocentric distance in the Galaxy (Wielen et al. 1992). Giant molecular clouds do not seem to heat the galactic disk very effectively. But they might play a crucial role if the primary source of heating the disk are short lived density wavelets because these lead predominantly to an increase of the planar velocity components of

the stars. Molecular clouds are then required to build up the vertical velocity distribution by deflections of stellar orbits out of the galactic plane (Jenkins & Binney 1990). Unfortunately, the excitation rate of the density wavelets is at present not quantitatively known. The effects of satellite galaxies sinking into disk galaxies and depositing their orbital energy as kinetic energy of the disk stars are presently studied in great detail (Quinn et al. 1993; Athanassoula 1993). However the duration of a single accretion event is only of the order of 10^9 years so that the diffusion of stellar orbits would then depend to a large degree on the accretion rate of satellite galaxies which is presently discussed rather controversial.

One of the principal difficulties of the scenario of massive black holes as the main constituents of the dark galactic halo is that one would expect a considerable dynamical evolution of the core of the black hole distribution near the galactic center. Due to dynamical friction such black holes spiral towards the galactic center and might eventually coalesce there. This accretion is balanced by "loss cone" depletion of stars near the galactic center due to the strong fluctuations of the gravitational field induced by the black holes and by "slingshot" effects of the black holes among themselves ejecting black holes from the core region. Hut & Rees (1993) argue that the latter mechanism might be not as effective as was previously assumed and conclude that about 100 of such massive black holes will reside close to the galactic center which is in clear contradiction to the observed kinematics of stars and interstellar gas near the galactic center. However, the interplay of the various dynamical effects is rather subtle and more refined numerical simulations of the dynamical evolution of the inner part of the black hole distribution are required in order to reach a more definitive conclusion.

References

Athanassoula E., 1993, in preparation
Edvardson B., Andersen J., Gustafsson B., Lambert D.L., Nissen P.E., Tomkin J., 1993, in preparation
Freeman K., 1991, in: Dynamics of Disk Galaxies, B. Sundelius (ed.), Göteborg University, Göteborg, p. 15
Fuchs B., 1980, PhD thesis, Univ. of Kiel
Grenon M., 1987, JA&A 8, 123
Hut P., Rees, M., 1993, MNRAS, in press
Jenkins A., Binney J., 1990, MNRAS 245, 305
Lacey C.G., 1984, MNRAS 208, 687
Lacey C.G., Ostriker J.P., 1985, ApJ 299, 633
Quinn P.J., Hernquist L., Fullagar D.P., 1993, ApJ 403, 74
Sellwood J.A., Carlberg R.G., 1984, ApJ 282, 61
Spitzer L., Schwarzschild M., 1951, ApJ 114, 385
Spitzer L., Schwarzschild M., 1953, ApJ 118, 106
Tóth G., Ostriker J.P., 1992, ApJ 389, 5
Wielen R., 1977, A&A 60, 263

Wielen R., Fuchs B., 1983, in: Kinematics, Dynamics, and Structure of the Milky Way, W.L.H. Shuter (ed.), Reidel, Dordrecht, p. 81

Wielen R., Dettbarn C., Fuchs B., Jahreiß H., Radons G., 1992, IAU Symp. 149, The Stellar Populations of Galaxies, B. Barbuy & A. Renzini (eds.), Kluwer, Dordrecht, p. 81

2. Tools of Ergodic Theory

Ergodic Methods in Stellar Dynamics

V.G. GURZADYAN

Dept. of Theoretical Physics, Yerevan Physics Institute, 375036 Yerevan, Armenia

1. Introduction

Stellar dynamics is one of the few well-established disciplines which has continuously attracted theorists throughout the present century. Moreover in recent decades a huge army of computer simulators has also entered this field in a very active way. Given these tremendous efforts one would have hoped to have had more results than we have today. Indeed one cannot claim that we unambiguously understand, say, even the basics of the structure of spiral galaxies, or of their origin and further evolution.

In my opinion this situation is only demonstrating the difficulty of the problem of the dynamics of gravitating systems. Ironically, the main problem of stellar dynamics was shown to be "unsolvable" before stellar dynamics itself was established in the first decades of this century. By this I mean the proof by Poincaré in the 1890's of the non-integrability of the gravitating N-body problem for $N > 2$.

Numerous approaches have been used to attack the dynamics of stellar systems, including various modifications of statistical mechanical, kinetic, hydrodynamical, celestial mechanical, and other methods. The treatment from the standpoint of ergodic theory can be considered as another possibility. As understood first in [1,2][1], it seems to reveal some profound aspects of the dynamics of stellar systems.

Below we will outline the key points of this approach, confining ourselves mainly to its methodological aspects. We also mention some advantages of this technique, while unsolved problems are listed in [3]. A more detailed account of these methods in the context of astrophysical and especially cosmological problems is contained in our recent monograph [4].

First let us note that by the term "stellar system" we mean a system of N point-like objects interacting by Newtonian gravity, i.e., described by a

[1] That study was done in the spring of 1983 (preprint EPI-678(68), 1983); we then had to wait one year (!) to get permission to send the manuscript abroad to *Astronomy & Astrophysics*.

Hamiltonian

$$H(p,r)=\sum_{a=1}^{N}\sum_{i=1}^{3}\frac{p_{(a,i)}^2}{2m_a}+U(r)\,, \tag{1}$$

$$U(r)=-\sum_{a<b}\frac{Gm_am_b}{|r_{ab}|}\,,\qquad r_{ab}=r_a-r_b\,. \tag{2}$$

In principle the generalisation of at least some of the results is possible for non-point-like objects, but we will not consider that case here.

The main problem to be addressed by an ergodic theory approach is *to discover to which class of dynamical systems stellar systems belong and to reveal the meaning of abstractly formulated properties in the context of those physical systems.*

2. Dynamical Systems

To have a discussion which is as self-contained as possible, we briefly present the elements of ergodic theory concerning the classification of smooth dynamical systems by the degree of their statistical properties. A detailed account of these points can be found in books on ergodic theory (see e.g. [5,6]).

Saying that $(M,\mathcal{B},\mu,T)$ is a dynamical system, where M is a smooth manifold, $\mathcal{B}$ is a σ–algebra of measurable sets on M, and μ is a complete normalised measure on $\mathcal{B}$, we mean that T^t is a differentiable one-parameter group of diffeomorphisms (a flow) defined by vector field $\mathbf{v}$ (the phase velocity) on M

$$\mathbf{v}(x)=\frac{\mathrm{d}T^tx}{\mathrm{d}t}. \tag{3}$$

The classification of flows following from ergodic theory includes ergodic systems, systems with mixing, K-systems, Axiom-A systems, and Anosov systems.

In the case of *ergodic systems*, for any measurable invariant set A

$$T^tA=A=T^{-t}A, \tag{4}$$

its measure μ should have the values

$$\mu(A)=\begin{cases}0\\1\end{cases}, \tag{5}$$

For measure-preserving ergodic transformations the time-average almost everywhere equals the phase space average

$$\int_M f\,\mathrm{d}\mu=\lim_{t\to\infty}\frac{1}{t}\int_0^t f\left(T^{-\tau}(x)\right)\,\mathrm{d}\tau \tag{6}$$

and all trajectories of the system pass through almost all points of the manifold M. This property, being both a necessary and sufficient condition of ergodicity, is often considered as a definition of an ergodic system. Ergodicity is also defined on manifolds with infinite measure but it is a rather weak statistical property.

A stronger and therefore more interesting property for physical problems is the property of mixing. Different levels of mixing exist. A dynamical system is said to possess the property of *weak mixing* if $\forall f, g \in L^2$

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t \left[\int_M f(T^{-\tau}x) g \, \mathrm{d}\mu - \int_M f \, \mathrm{d}\mu \int_M g \, \mathrm{d}\mu \right]^2 \mathrm{d}\tau = 0 \, ; \tag{7}$$

the property of *mixing (strong mixing)* if

$$\lim_{t\to\infty} \int_M f(T^t x) g \, \mathrm{d}\mu = \int_M f \, \mathrm{d}\mu \int_M g \, \mathrm{d}\mu \, ; \tag{8}$$

and the property of *n-fold mixing* if

$$\lim_{t_1,\ldots,t_m\to\infty} \int_M f_0 f_1(T^{t_1}x) \ldots f_m(T^{t_1+\ldots+t_m}x) \, \mathrm{d}\mu = \prod_{i=0}^{m} \int_M f_i \, \mathrm{d}\mu \, . \tag{9}$$

These properties describe systems with increasingly stronger statistical properties in the sense that systems with mixing possess the property of weak mixing, and those with n-fold mixing also that of mixing and weak mixing but not vice versa.

Therefore systems with mixing are also ergodic ones. However, for the systems with mixing, as opposed to ergodic ones, a set $A \in \mathcal{B}$ evolves in such a way (preserving its measure and connection) that the measure of the part which intersects the set $B \in \mathcal{B}$ tends in time to be proportional to the measure of B

$$\lim_{t\to\infty} \frac{\mu(T^t A \bigcap B)}{\mu(A)} = \mu(B) \, . \tag{10}$$

This property implies the existence of a final state of measure μ to which

$$\mu_t(A) = \mu_0(T^t A), \quad A \in \mathcal{B} \tag{11}$$

tends continuously, so that

$$\lim_{t\to\infty} \int_M f \, \mathrm{d}\mu_t = \int_M f \, \mathrm{d}\mu \, . \tag{12}$$

It is natural to call the final state given by μ *equilibrium* and the process of tending to that state *relaxation.*

Thus we see that only with systems with the mixing property can one speak about a relaxation process and an equilibrium state, which is not possible for systems which are only ergodic. This fact is clearly seen from the difference between the limit

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t \mu\left(T^{-\tau}A\cap B\right)\mathrm{d}\tau=\mu(A)\,\mu(B)\ , \tag{13}$$

for ergodic systems and the limit (10) for mixing ones; the first limit is said to be convergent *in the Cesaro sense*, while in the second case one has real convergence, i.e., for ergodic systems the initial fluctuations tend to zero only in the time-average, while for mixing systems their absolute value decreases as well.

The weak mixing is not sufficient as well, as follows from the limit

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t \left|\mu\left(T^{-\tau}A\cap B\right)-\mu(A)\,\mu(B)\right|\mathrm{d}\tau=0, \tag{14}$$

implying that T^tA becomes independent of the set B only if some parts of the trajectory are not taken into account.

Even stronger statistical properties characterise K-mixing/Kolmogorov systems for which the following limit

$$\lim_{t\to\infty}\sup\left|\mu(A\cap B)-\mu(A)\,\mu(B)\right|=0 \tag{15}$$

exists, where the upper limit is taken for the smallest σ-algebra containing sets $T^{t_0}A$ for $t<t_0$. Kolmogorov systems possess n-fold mixing with arbitrary n. The strongest statistical properties characterise hyperbolic (Axiom-A), Anosov, and Bernoulli systems.

A dynamical system f^t is said to be of Anosov type if for its all trajectories $\{f^t\}$ there exist subspaces $E^s(f^t(x))$, $E^u(f^t(x))$ of the tangential space $TM_{f^t(x)}$, and numbers $C>0$, $\lambda>0$, such that

$$\begin{aligned} TM_{f^t(x)} &= E^s\left(f^t(x)\right)\oplus E^u\left(f^t(x)\right)\ ,\\ \mathrm{d}f^\tau E^s\left(f^t(x)\right) &= E^s\left(f^{t+\tau}(x)\right)\ ,\\ \mathrm{d}f^\tau E^u\left(f^t(x)\right) &= E^u\left(f^{t+\tau}(x)\right)\ , \end{aligned} \tag{16}$$

and for all $t>0$, one has

$$\begin{aligned} &\|\mathrm{d}f^tv\|\le C\,\mathrm{e}^{-\lambda t}\,\|v\|\ , \quad v\in E^s\ ;\\ &\|\mathrm{d}f^tv\|\ge C^{-1}\mathrm{e}^{\lambda t}\,\|v\|\ , \quad v\in E^u\ . \end{aligned} \tag{17}$$

The subspaces E^s and E^u are called the stable (converging) and unstable (expanding) subspaces. Anosov systems are a typical subclass of hyperbolic systems. Geodesic flow on a compact manifold with negative constant curvature, studied long ago by Hadamard, Hopf and Hedlund, is an example of an Anosov system [7]. If the systems with mixing can tend to equilibrium by any law (e.g. polynomial), then Anosov and Axiom-A (hyperbolic) systems tend to that state exponentially.

The problem of distinguishing different features of dynamical systems, and the formulation of corresponding characterising criteria is one of the

central ones in ergodic theory. Many efforts in this direction have concerned the study of the so-called spectral properties of dynamical systems, and after 1958, when Kolmogorov discovered the new metric invariant, the entropy, also the entropy theory of dynamical systems.

Consider the entropy of a splitting ξ_i of the measurable manifold M

$$H(\xi) = \sum_{i=1}^{d} \mu(\xi_i) \ln(\xi_i) , \tag{18}$$

where $\xi_i \in \mathcal{B}$ and

$$\xi_i \bigcap_{i \neq j} \xi_j = \emptyset , \qquad \bigcup_{i=1}^{d} \xi_i = M . \tag{19}$$

Then the Kolmogorov-Sinai (KS) entropy h is the limit

$$h(f) = \sup \lim_{n \to \infty} \frac{1}{n} H(\xi^n) , \qquad \text{where} \qquad \xi^n = \bigvee_{j=0}^{n-1} f^{-j} \xi , \tag{20}$$

and the upper limit is taken over all measurable splittings.

Dynamical systems with positive KS-entropy $h > 0$ are usually called *chaotic*, while those with $h = 0$ are called *regular* ones. In particular, Anosov and Kolmogorov systems, which are typical systems with mixing, have positive KS-entropy $h > 0$, while most of only ergodic ones have $h = 0$. Therefore the latter are not considered to be chaotic according to this definition.

For the above mentioned geodesic flows on spaces with constant negative curvature $R < 0$, the KS-entropy equals

$$h = \sqrt{-R}. \tag{21}$$

The KS-entropy is related to the Lyapunov characteristic exponents λ_i via the Pesin formula:

$$h(f) = \int_M \sum_{\lambda_i(x) > 0} \lambda_i(x) \, d\mu(x); \tag{22}$$

we see that a system with at least one non-zero Lyapunov exponent has positive KS-entropy.

Finally let us mention another important characteristic of dynamical systems, the correlation function, defined by

$$b_{g,g'}(t) = \int_M g(f^t x) \, g'(x) \, d\mu - \int_M g(x) \, g'(x) \, d\mu. \tag{23}$$

Although at present estimates of the correlation functions (including numerical results on some billiards) exist only for a few dynamical systems, for

Anosov systems it has been shown that the correlation functions decay exponentially, i.e., $\exists \alpha_{g,g'}, \beta, t > 0$ so that

$$|b_{g,g'}(t)| \leq \alpha_{g,g'} \exp(-\beta t) , \qquad \text{where} \qquad \beta \simeq h(f) . \tag{24}$$

3. Geodesics

As mentioned above geodesic flows on spaces with negative curvature were long ago an object of study in ergodic theory. In view of the method existing in classical mechanics even earlier, known as the Maupertuis principle [8], enabling one to represent the phase flow of a Hamiltonian system as a geodesic flow on some Riemannian manifold, one has a very real possibility of treating the gravitating N-body problem with concepts of ergodic theory. Just this idea was the basis of the original study of stellar dynamics in papers [1, 2]. In physical problems this idea, as far as we know, was first used by Krylov [9].

By means of the Maupertuis principle the Hamiltonian equations

$$\dot{r}^{\mu} = \frac{\partial H}{\partial p_{\mu}}, \qquad \dot{p}_{\mu} = -\frac{\partial H}{\partial r^{\mu}} , \tag{25}$$

are reduced to the geodesic equation

$$\nabla_u u = 0 \tag{26}$$

on the region of configuration space

$$M = \{W = E - V(r) > 0\}$$

with the Riemannian metric

$$\mathrm{d}s^2 = [E - V(r_{1,1}, \ldots, r_{N,3})] \sum_{a=1}^{N} \sum_{i=1}^{3} (\mathrm{d}r_{a,i})^2 , \tag{27}$$

where E is the total energy of the system. The condition of conservation of the total energy of the system

$$H(p, q) = E \tag{28}$$

is equivalent to the condition on the velocity associated with the geodesic

$$||u|| = 1 , \tag{29}$$

while the affine parameter along the geodesic is determined by

$$\mathrm{d}s = \sqrt{2}\big(E - V(r)\big)\mathrm{d}t . \tag{30}$$

The statistical properties of the geodesic flow, particularly their dependence on the curvature of the manifold as mentioned above, are seen from the geodesic deviation equation

$$\nabla_u \nabla_u n + \mathrm{Riem}(n, u)\, u = 0 \,. \tag{31}$$

This equation, known as the Jacobi (or Jacobi–Levi-Civita) equation, can be derived by direct differentiation of the geodesic equation along the deviation vector **n** and using the following expressions involving the Riemann tensor

$$\begin{aligned} &\mathrm{Riem}(n, u)\, u = \Big([\nabla_n, \nabla_u] - \nabla_{[n,\,u]}\Big)\, u = [\nabla_n, \nabla_u]\, u \,, \\ &\nabla_n u - \nabla_u n = [n, u] = 0 \,. \end{aligned} \tag{32}$$

For a vector field satisfying the condition of orthogonality

$$\langle n, u \rangle = 0 \,, \tag{33}$$

the Jacobi equation may be expressed in the form

$$\frac{d^2 \|n\|^2}{ds^2} = -2K_{u,n}\|n\|^2 + 2\|\nabla_u n\|^2 \,, \tag{34}$$

where $K_{u,n}$ denotes so-called two-dimensional curvature

$$K_{u,n} = \frac{\langle \mathrm{Riem}(n, u)u, n\rangle}{\|u\|^2\|n\|^2 - \langle u, n\rangle^2} = \frac{\langle \mathrm{Riem}(n, u)\, u, n\rangle}{\|n\|^2}. \tag{35}$$

One sees that the Jacobi equation describes geodesics which diverge no slower than by an exponential law

$$\|n(s)\| \geq \frac{1}{2}\|n(0)\| \exp\left(\sqrt{-2k}s\right), \qquad s > 0 \tag{36}$$

if the following condition is fulfilled

$$k = \max_{u,n}\{K_{u,n}\} < 0 \,. \tag{37}$$

Geodesic flow in this case represents an Anosov system with the strongest statistical properties as mentioned above. *Thus the strong negativity of the two-dimensional curvature is an important criterion linking the geometrical properties of the Riemannian manifold with the statistical properties of the geodesic flow.*

Before discussing the N-body problem we mention another case of transferring the problem of motion in real space to the study of the properties of a Riemannian space. The problem concerns the study of the hydrodynamical equations of motion by means of the group of diffeomorphisms $\mathrm{Sdiff}(D)$ of a Riemannian manifold D with a one-sided invariant metric as considered by Arnold [10] (and later by a number of authors).

Examining the Lie algebra of all vector fields with zero divergence on the two-dimensional torus T^2, it can be shown that the kinetic energy of the element of the moving fluid induces a right-invariant Riemannian metric on $\mathrm{Sdiff}(T^2)$. The principle of least action, which determines the motion of an incompressible fluid in terms of the geodesics of this metric, plays the role of the Maupertuis principle. In the context of stellar dynamics this method has been used in [11].

4. Curvature

Turning now to the gravitating N-body problem, in view of the above considerations one should first calculate the two-dimensional curvature $K_{u,n}$ for the Hamiltonian (1). This curvature was calculated and analysed in detail in [1, 2] and has the form

$$\begin{aligned}\mathrm{Riem}_{\mu\lambda\nu\rho} = &-\frac{1}{2W}\left[g_{\mu\nu}W_{\lambda\rho} + g_{\lambda\rho}W_{\mu\nu} - g_{\mu\rho}W_{\lambda\nu} - g_{\nu\rho}W_{\mu\lambda}\right] \\ &-\frac{3}{4W^2}\left[(g_{\mu\lambda}W_{\nu} - g_{\mu\nu}W_{\lambda})W_{\rho} + (g_{\nu\rho}W_{\lambda} - g_{\lambda\rho}W_{\nu})W_{\mu}\right] \\ &-\frac{1}{4W^2}\left[g_{\mu\nu}g_{\lambda\rho} - g_{\mu\lambda}g_{\nu\rho}\right]\|\partial W\|^2 ,\end{aligned} \tag{38}$$

where

$$g_{\mu\nu} = W\delta_{\mu\nu} , \qquad W_{\mu} = \frac{\partial W}{\partial r^{\mu}} , \qquad W_{\mu\nu} = \frac{\partial^2 W}{\partial r^{\mu}\partial r^{\nu}} . \tag{39}$$

As expected the quantity $K_{u,n}$ is sign-indefinite, taking both negative and positive values in different regions of phase space.

However, it was shown that for configurations with a spherical distribution of coordinates and velocities, two-dimensional curvature is determined by the scalar curvature R

$$K_{u,n} = \frac{R}{3N(3N-1)} \tag{40}$$

and is negative since the scalar curvature is negative for $N > 2$:

$$R = -\frac{3N(3N-1)}{W^3}\left(\frac{1}{4} - \frac{1}{2N}\right)(\nabla W)^2 - (3N-1)\frac{\Delta W}{W^2} < 0 . \tag{41}$$

From here, using results on the properties of geodesic flows on spaces with negative curvature as briefly described above, one can conclude that N-body systems situated in the region of phase space with negative scalar curvature have positive KS-entropy (determined by the value of the scalar curvature) and thus possess strong mixing properties peculiar at least to Kolmogorov systems. The KS-entropy, i.e., the scalar curvature, therefore determines the

rate of exponential decay of the correlation functions and the characteristic time scale for approaching an equilibrium state. If such a state exists one can speak about a *relaxation* process.

An important property of systems with exponential mixing is their *structural stability (coarseness)*. In our case this means that some global properties of spherical systems, such as the number density of stars and their velocity dispersion, do not depend exactly on the initial conditions of the system, namely, on the coordinates and velocities of each star. In contrast some other properties depend on them strongly, e.g., two stars close at one moment later could be highly separated by the smallest change of those conditions.

5. Hierarchy of Characteristic Time Scales of the Stellar System

Arriving at the concept of the relaxation time of a stellar system we cannot avoid discussing the role of various characteristic time scales in the dynamics and evolution of the stellar system. In fact one can speak of a hierarchy of time scales since they are not only associated with different aspects and epochs of the evolution of the stellar system, but can also vary within rather different quantitative ranges for a single system.

The sequence of these scales can be represented as follows:

1.	τ_{dyn}	non-stationary phase	N-body
2.	$\tau_{N\text{-body}}$	stationary	N-body
3.	τ_{binary}	stationary	2-body
4.	τ_{evol}	evolutionary effects	N-body

Which of these characteristic time scales can be designated as *relaxation times*?

The dynamical time scale τ_{dyn}, as known since the classic result of Lynden-Bell, describes essentially non-stationary systems (violent relaxation) tending to a coarse-grained equilibrium. As we have seen above the linkage

curvature – geodesics – KS-entropy – mixing – equilibrium

enables us to associate $\tau_{N\text{-body}}$ with the relaxation process leading to some equilibrium state (fine-grained state); more precisely it is the time scale for returning to the initial state of a perturbed quasi-stationary system. Evidently, both the Lynden-Bell violent relaxation time and this time scale take into account the gravitational interaction of all N particles of the system.

Using some reasonable physical assumptions, an explicit formula for the characteristic *relaxation time scale* has been derived in [1, 2] which can be attributed to spherical systems with some average values of the number density of particles (stars) and of the velocity dispersion. The formula so derived does not coincide with that of the well-known binary (Chandrasekhar) relaxation time scale evaluated for the same quantities and seems to be supported by observational data on globular clusters and elliptical galaxies [12, 13, 14].

In other words the contribution of two-body interactions for typical stellar systems seems not to be the dominant one among N-body interactions. Indeed, it is hard to imagine say that the dynamics of a particle from a typical 3-body system would not be changed if the contribution of one of particles were neglected, or of the two particles from the 4-body system, etc. At the same time the mathematical background we have used shows that when one speaks of the concept of a *relaxation time scale* for a stellar system one should understand clearly the approximations made in its definition, and hence its limited physical meaning.

Indeed, from the sign-indefiniteness of the two-dimensional curvature it follows that no universal function

$$\tau : (\text{all } N\text{-body systems}) \to R_+$$

exists and hence no unique relaxation time scale can exist for all N-body gravitating systems. Moreover for the same system different physical quantities can relax to their equilibrium values in different times, others cannot relax at all, etc.

By selecting those systems situated in the region of phase space with negative two-dimensional curvature, we thus limit ourselves to considering only the epoch of evolution of the stellar system during which the latter keeps its quasi-stationary spherical structure. In this local (in time) problem we ignore the radical modification of the system during the evolution, thus neglecting the main difficulty: the non-compactness of the phase space of the system.

The main motivation for this assumption is the fact that the time scale τ_{evol} of radical modifications of the system for real stellar systems greatly exceeds the time scales we discussed above. The evolution driving effects include both dynamical effects, such as evaporation of stars from the system and core collapse, as well as other effects (mass loss by stars, formation of binaries via tidal two-body interactions, star collisions, etc). In the latter case the problem is evidently not a purely stellar dynamical one.

These considerations concern mainly spherical systems. The situation with disk systems, for example, is rather different. In [11], while studying the stellar motion in the disk of the Galaxy in the hydrodynamical approximation, namely considering the Lie algebra $\text{Sdiff}(T^2)$ with the current function ξ of a velocity field

$$v = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \text{grad}\, \xi \,, \tag{42}$$

the exponential instability of the motion was established, while the initial velocity field was shown to remain stable. The characteristic time of the instability is less than the period of rotation of the disk. This example, incidentally, indicates that not every exponential instability must be related

to relaxation. Another point of view on disk-like stellar configurations as dynamical systems with non-negative curvature is developed in [15, 16].

6. Ricci Curvature

The absence of rigorous results concerning many-dimensional nonlinear systems with sign-indefinite curvature, together with doubts about the possibility of having some distinguishing criteria of practical interest in the general case, force one to look for average geometrical characteristics.

A criterion based on an averaged quantity, the Ricci curvature, was introduced in [17] as a measure of the relative instability of many-dimensional dynamical systems. The method has been already applied to the dynamics of stellar systems, to a system situated in a regular field (the regular field was shown to increase the degree of instability of the nonlinear system) [18], and to plasma configurations [19].

The idea is as follows. Average the Jacobi equation over the geodesic deviation vector

$$\frac{\mathrm{d}^2 z}{\mathrm{d}s^2} = \frac{1}{3N} r_u(s) + \langle \|\nabla_u n\|^2 \rangle \,, \quad \text{where} \quad n = z\hat{n} \,, \quad \|\hat{n}\|^2 = 1 \,, \tag{43}$$

and $r_u(s)$ denotes the Ricci curvature in the direction of the velocity of the geodesic (Ric is the Ricci tensor)

$$r_u(s) = \frac{\mathrm{Ric}(u,u)}{u^2} = \sum_{\mu=1}^{3N-1} K_{n_\mu, u}(s) \,, \qquad (n_\mu \perp n_\nu, n_\mu \perp u) \,. \tag{44}$$

The Ricci tensor has the expression

$$\begin{aligned} \mathrm{Ric}_{\lambda\rho} = &- \frac{1}{2W} \left[\Delta W g_{\lambda\rho} + (3N-2) W_{\lambda\rho} \right] \\ &+ \frac{3}{4W^2} \left[(3N-2) W_\lambda W_\rho \right] - \left[\frac{3}{4W^2} - \frac{(3N-1)}{4W^2} \right] g_{\lambda\rho} \|dW\|^2 \,. \end{aligned} \tag{45}$$

The criterion of relative instability defined in [17] reads:
the more unstable of two systems is the one with smaller negative r

$$r = \frac{1}{3N} \inf_{0 \le s \le s_*} [r_u(s)] \,, \qquad r < 0 \tag{46}$$

within a given interval $0 \le s \le s_*$, *i.e., this system should be unstable with a higher probability in the same interval.*

Recalling that Riem depends on (x, u, n), Ric depends on (x, u), and R depends on (x), one can see that while the scalar curvature R cannot contain any worthwhile information about a typical system since it does not depend

on the velocities, Ric does contain them and therefore can be a characteristic quantity for typical (averaged) systems.

Numerical exploitation of the Ricci criterion of relative instability for the different models of stellar systems has shown that, e.g., a spherical system with a central mass is more unstable than a homogeneous one, spherical systems are more instable than disk-like ones, etc [17]. Thus, a classification of different configurations of stellar systems by increasing degree of statistical properties has been obtained.

7. Conclusions

What are the main advantages of the present treatment of stellar dynamics from the point of view of theory of dynamical systems?

The **first** major gain seems to be the precise definitions and hence the unique meaning of the terminology and notation often used in stellar dynamics, such as *relaxation, instability, chaos*, etc.

In particular, we have seen that the term *relaxation* cannot necessarily be attributed only to processes which involve exchange of energy between particles. The classical problem of the relaxation of an ideal gas is an illustration of this fact.

Another example is the term *chaos*: only systems possessing at least the property of mixing, and not purely ergodic ones, can be called chaotic, e.g., such terms as *ergodic chaos* should be considered as having no meaning. Spherical stellar systems can thus be designated as chaotic systems.

The **second** advantage is the possibility of obtaining deeper insight into the fundamental properties of N-body systems and the discovery of underlying relationships between them.

The relations between the negativity of the two-dimensional curvature determined by the Newtonian interaction potential, the exponential instability of phase trajectories and relaxation driving effects for spherical stellar systems, the properties of Lie groups, and the conclusion regarding the exponential instability of flow (but not relaxation) in the consideration of disk systems, are examples of this fact.

The absence of a universal relaxation time scale indicates that the attempts to estimate relaxation time scales which describe the observed states of galaxies and star clusters by using numerical simulations of 100 or 1000 or even more particle models of N-body systems with a somehow softened potential cannot make much sense. In particular, the exponential growth of errors in gravitating N-body systems as first found by Miller [20] cannot be directly related to relaxation time scales since the same effect can be seen, for example, in the integrable 2-body Kepler problem (for the discussion of N-body numerical simulations see [21,22]).

The **third** advantage is the development of new methods and criteria which can be used both in analytical and especially in computer studies of N-body systems.

Studies already completed indicate that the Ricci curvature method can be a useful tool for numerically investigating the local (in time) instability properties not only of stellar configurations, but also of plasma and other many-dimensional systems.

Finally, this approach offers more unsolved problems, some of which are listed in [3], than those which are presently solvable. I think this can be considered as another advantage rather than as a weak point of this technique.

Acknowledgements. I sincerely thank Armen Kocharyan for valuable discussions and a number of helpful comments. I am very much obliged to George Savvidy, my collaborator and coauthor of an important paper.

References

[1] Gurzadyan V.G., Savvidy G.K., 1984, Doklady AN SSSR, 277, 69
[2] Gurzadyan V.G., Savvidy G.K., 1986, A&A, 160, 203
[3] Gurzadyan V.G., *10 Problems* (this volume)
[4] Gurzadyan V.G., Kocharyan A.A., 1993, Paradigms of The Large-Scale Universe, Gordon and Breach (in press)
[5] Kornfeld I.P., Sinai Ya.G., Fomin S.V., 1980, Ergodic Theory, Nauka, Moscow
[6] Dynamical Systems. Modern Problems in Mathematics, 1985, vol.2, Ed. Sinai Ya.G., VINITI Publ., Moscow
[7] Anosov D.V., 1967, *Geodesic Flows on Closed Riemannian Spaces with Negative Curvature*, Comm. MIAN, vol.90
[8] Arnold V.I., 1989, Mathematical Methods of Classical Mechanics, Nauka, Moscow
[9] Krylov N.S., 1950, Studies on Foundation of Statistical Mechanics, Publ. AN SSSR, Leningrad
[10] Arnold V.I., 1966, Annales de l'Institut Fourier, XVI, 319
[11] Gurzadyan V.G., Kocharyan A.A., 1988, A&A, 205, 93
[12] Vesperini E., 1992, Europhys.Lett, 17, 661
[13] Vesperini E., 1992, A&A, 266, 215
[14] Pucacco G., 1992, A&A, 259,473
[15] Gurzadyan V.G., Kocharyan A.A., 1986, Doklady AN SSSR, 287, 813
[16] Gurzadyan V.G., Kocharyan A.A., 1987, Ap&SS, 133, 253
[17] Gurzadyan V.G., Kocharyan A.A., 1987, Ap&SS, 135, 307
[18] Gurzadyan V.G., Kocharyan A.A., 1988, Doklady AN SSSR, 301, 323
[19] Cosentino S., Gurzadyan V.G., Kocharyan A.A., Ruffini R., 1989, Preprint University of Rome "La Sapienza"
[20] Miller R.H., 1964, Ap.J., 140, 250
[21] Gurzadyan V.G., 1992, in: "Dynamics of Globular Clusters", Proceed. Ivan King Workshop, Berkeley
[22] Gurzadyan V.G., Kocharyan A.A., (this volume)

On a Notion of Weak Stability and Its Relevance for Celestial Mechanics and Molecular Dynamics

Luigi GALGANI & Antonio GIORGILLI

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

Abstract. The stability problem in Hamiltonian dynamics is discussed in the light of Nekhoroshev's theorem. This guarantees a form of weak stability, namely referred to finite (rather than infinite) times. Applications are discussed for the restricted problem of three bodies and for the problem of energy equipartition in statistical mechanics.

1. Introduction

The problem of stability in Hamiltonian dynamical systems is of fundamental importance, and every improvement in its mathematical formalization is consequently of great interest. After the years 1954–1963 a new way of dealing with it was afforded by the celebrated Kolmogorov (or Kolmogorov–Arnold–Moser) theorem, which could guarantee stability in a certain weak sense, namely for a set of initial data of large measure (while a full result could be obtained only for systems of two degrees of freedom). In the present lecture we intend to make a little propaganda for another approach, which usually goes under the name of Nekhoroshev [1,2], a pupil of Arnold who published the relevant paper in the year 1977; apparently, it occurred to Benettin, Gallavotti, and the authors to be the first ones to become fully acquainted with such a theorem in the West [3,4]; for applications to statistical mechanics and to the realization of holonomic constraints, see [5,6]. In fact, it then became clear that such a method was already known to Moser [7] and especially to Littlewood [8], who used it also for applications to celestial mechanics; moreover, the corresponding philosophy seems to go back to Boltzmann [9,10] and Jeans [11,12], dating to almost a century ago. Furthermore, some similar results are well known to people working in applied physics, such as plasma physics [13,14] and molecular dynamics [15,16]. For what concerns the references, one can obviously refer to the original papers; we also like to mention three papers (see [17], [18] and [19]), where a certain effort was made to be pedagogical. In the present, very short, review, we shall instead just give some information at a crude informal level.

2. Stability in Nekhoroshev's Sense, and an Application to the Problem of Three Bodies

Let us start with the main point, namely the definition of stability itself, considering the paradigmatic case of an equilibrium point for a differential equation in $\mathbf{R}^n$. According to the usual definition, an equilibrium point is said to be stable if for any neighborhood U of it there exists another neighborhood V such that all orbits starting in V remain in U for *all* times. This is certainly a clearcut definition, but it is questionable whether it is the interesting one. Two points come immediately to one's mind: (i) the problem of the stability basin (to have estimates on physically relevant invariant neighborhoods of the equilibrium) (ii) the problem of the stability time (in many applications, it is sufficient to be insured that initial data close to the equilibrium give rise to orbits remaining in an interesting neighborhood of it for a *sufficiently long* time, independently of what will occur for larger times). For example, in applications to celestial mechanics, stability over a *finite* time such as the estimated age of the universe might be in practice sufficient. As Littlewood said, this is not the eternity, but a considerable slice of it.

In fact, such two problems are strictly connected. This is well understood if one just recalls the standard existence and uniqueness theorem for ordinary differential equations (in normal form), because it guarantees continuous dependence on initial data (and parameters). Indeed from such a theorem there follows that, given any neighborhood U of the equilibrium and a time T, there correspondingly exists a neighborhood V such that all orbits starting in V will remain in U for all times t with $|t| < T$. But this is trivial, and by the way is true also for unstable equilibria. As a consequence, if one looks for a large T, then, in general, V must be ridiculously small with respect to U; if instead one wants U and V to be, in some sense, of the same order of magnitude, then T turns out to be ridiculously small. As a trivial example, consider the linear equation $\dot{x} = x$.

The Nekhoroshev's type approach allows instead to have much stronger results in many interesting cases. As a typical example, consider the case of an equilibrium of an analytic Hamiltonian system with n degrees of freedom, and assume that the eigenvalues $i\omega_l$, $(l = 1, \ldots, n)$ of the linearized system (i) are pure imaginary (i.e., the equilibrium is elliptic), and (ii) satisfy the diophantine condition $|k \cdot \omega| > \gamma |k|^{-\tau}$ for all non-vanishing $k \in \mathbf{Z}^n$, and for some constants $\gamma > 0$ and $\tau > n - 1$. Then one can prove (see [20]) that there exists a critical distance $\bar{d}$ from the equilibrium, such that for all initial data with distance $d < \bar{d}$ the time required for the orbit to reach distance $2d$ increases exponentially with $(1/d)^{1/(\tau+1)}$. Thus, having fixed a time T, one gets stability regions within T which might be rather large. In addition, a more accurate analysis shows that the method allows somehow to build up explicitly a natural domain within which stability up to very large times is

guaranteed. A similar result can be given also in the case of a reversible (non necessarily Hamiltonian) system; see [21].

The first application to a physically relevant system concerns the restricted problem of three bodies, and was given by Giorgilli et al.[22]; some improvements were then added in [23] and in [24]. One considers the Lagrangian equilibrium point L_4 of the Sun–Jupiter system. It is known that such an equilibrium is linearly stable, but essentially no results do exist for the nonlinear system in three dimensions (although perpetual stability can be inferred by KAM theorem in the planar circular case). The result found is that there exists a significant domain of weak stability in the sense described above (namely no escape is guaranteed from a domain of double size), over a time as large as the estimated age of the universe: precisely, the size of the domain is larger than 4×10^{-5} times the distance of Jupiter from L_4. This is still below the distance of the known asteroids, but already of a realistic order of magnitude. An analogous situation obtains for the Lagrangian L_4 point of the Earth–Moon system, where the stability region in the above sense is estimated to have a radius larger than 2×10^{-4} times the distance of the Moon from the Lagrangian point. Particularly interesting are the figures of the paper [24], where the stability time is reported versus the distance from the Lagrangian point. Coming from the region of high distances, one sees the time to increase slowly as expected from the existence and uniqueness theorem, and then to rise sharply, in a qualitatively striking way, below the critical distance, somehow simulating a practically infinite stability time.

3. The Idea of the Proof

Avoiding to enter into the details of the proof, we shall only mention here the main ingredient, referring for simplicity to a Hamiltonian perturbation of a system of harmonic oscillators. So, one considers a Hamiltonian system of the form

$$H(p,q,\varepsilon) = h(p) + \varepsilon f(p,q,\varepsilon) \ , \tag{1}$$

where (p,q) are action-angle variables, and H is assumed to be an analytic function of all the variables; here, $h(p) = \sum_j \omega_j p_j$ is the integrable part, f is the perturbation, and ε a small parameter. In this case $\dot{p}$ is of order ε, which corresponds to stability times of order ε^{-1}. Assuming that the frequencies ω are non-resonant, ordinary perturbation theory looks for canonical transformations such that in the new variables (P,Q) the Hamiltonian takes a form of the type

$$H'(P,Q,\varepsilon) = h(P) + \varepsilon g(P,\varepsilon) + \varepsilon^r f'(P,Q,\varepsilon) \ , \tag{2}$$

with suitable g and f', and a positive integer r (perturbative order); so one has that $\dot{P}$ is of order ε^r, and formally this corresponds to a stability time of order ε^{-r}.

In fact this is only formal, because in general one provides estimates neither of the domains (in the p variables) where the transformations are defined, nor of the size of the new perturbation f'. The essential point in Nekhoroshev's approach is a careful control of such estimates. First of all, the domain has to be determined in such a way that resonances due to $g(P)$ should be avoided; moreover, due to the diophantine condition quoted above, the size of f' turns out to grow typically as $(r!)^n$. As a consequence of the latter estimate one has thus that, at order r , the perturbation (or remainder) is of typical size $(r!)^n \varepsilon^r$. A standard argument in the theory of asymptotic expansions then leads the to a choice of an optimal order $r_{\text{opt}}(\varepsilon)$ as the one minimizing the remainder, and one finds $r_{\text{opt}}(\varepsilon) \sim (1/\varepsilon)^{1/n}$, which leads to the estimate $\dot{P} \sim \exp[-(1/\varepsilon)^{1/n}]$. This is the way the exponential stability time comes about.

4. Application to the Foundations of Statistical Mechanics

So the main rationale of Nekhoroshev's method is to renounce to have information for all times (the eternity), and to ask questions only up to finite, very large, times (a considerable slice of the eternity). In fact such an attitude goes back to Boltzmann, who made use of it in discussing the paradox of reversibility and of recurrence in statistical mechanics. As is rather well known, Boltzmann remarked that for large systems the time required for practical recurrence (or in general for significant fluctuations away from statistical equilibrium) would be much larger than the age of the universe. What is usually not known, instead, is that a similar *escamotage* was considered by Boltzmann also in connection with the problem of equipartition [9,10].

Considering for simplicity the case of a system of (weakly perturbed) harmonic oscillators, the equipartition principle states that every frequency would get, in the average, the same energy. In fact, this should be true only by considering time averages over a sufficiently long time (relaxation time); but the size of the relaxation time is usually not discussed (see [25]). Anyhow, the suggestion of Boltzmann is that such a time would increase with the frequency, becoming of the order of days or centuries in some meaningful cases. Shortly later, Jeans suggested an exponential law. If this were true, in classical mechanics there would be no equipartition paradox (in the weak sense). And this would be very strange, because classical mechanics was always supposed to fail just in this connection, namely exactly where quantum mechanics arose. So it is clear that this is a quite relevant question.

The problem was raised in more recent times (1954) by Fermi–Pasta–Ulam [26], with their celebrated study on the one–dimensional solid (for a review see [27]). Izrailev and Chirikov [28] were the first to pick up again the problem. Followed by J. Ford, they put forward the suggestion that in

the thermodynamic limit (number n of degrees of freedom tending to infinity, with finite specific energy), equipartition would be recovered. On the basis of the numerical computations of Bocchieri, Loinger and Scotti[29], it was suggested instead by Cercignani, Galgani and Scotti[30,31] that equipartition could obtain only for specific energies above a critical non-vanishing threshold. There were a lot of numerical computations, and a few analytical studies. All analytical indications seemed to be against an extrapolation of KAM methods to the thermodynamic limit. We then became aware of the Nekhoroshev point of view, and the original proposal of Boltzmann and Jeans were rediscovered[32].

The most delicate point for an application of Nekhoroshev's–type results to statistical mechanics is that in Nekhoroshev's estimates one usually gets stability times of the order

$$T_* e^{(\omega/\omega_*)^a} , \tag{3}$$

where ω is the frequency involved, T_* and ω_* are constants depending on n, while the constant a is typically of the order of $1/n$; just because of this estimate of the constant a, it is thus clear that the exponential would disappear in the thermodynamic limit. However, a first breakthrough was obtained when it was realized that there exist physically interesting models of n degrees of freedom for which one has indeed $a = 1$. The typical case is that of a system which splits into two subsystems characterized by well separated frequencies, with the additional property that the high frequency subsystem is completely resonant; the simplest model of this type is a one dimensional system of identical diatomic molecules, interacting through an analytic intermolecular potential. For systems of this type one can renounce to control the energy exchange among the single oscillators, being satisfied with a control of the exchange of energy between the two subsystems. In such a way one gets exactly $a = 1$. A numerical investigation on the system of diatomic molecules was given in ref. [33], while an analytical result was given in [5].

Having eliminated the worst dependence of the estimates on n for models of such a type, there still remains to discuss the dependence of the other constants T_* and ω_* on n; the simplest available estimates give that they could decrease to zero as inverse powers of n, but it is still an open problem whether such estimates are optimal.

A numerical investigation was then made[34] for a modification of the FPU model, in which the equal masses were replaced by alternating heavy and light masses. Indeed, it is well known that for such a modified model one has a splitting of the frequency spectrum into two well separated branches, called the acoustic and the optical branches, characterized by low and high frequencies respectively. The numerical indication was that for such a model, for initial data with *fixed specific energy* E/n, the constant ω_* is independent of n, while T_* should decrease to zero at most as $1/\ln n$. Thus, the energy exchange between the two subsystems could remain small for exponentially large times even for macroscopic systems with $n \sim 10^{23}$.

More recently an analytical result along these lines (control of energy exchange between subsystems) was proved [35] for a rather large class of systems, which includes the modified FPU model. In particular, for the latter model one proves that the energy exchange between the two subsystems is small for exponentially large times, with constants independent of n, provided the *total* energy E is fixed, which corresponds to vanishing specific energy E/n. Whether such a result could be improved to a control of the energy exchange for finite *specific* energy is not known; we hope to be able to prove it at least in a statistical sense, namely for a set of initial data of large measure.

5. Final Remarks

We hope we were able to show how the applications of the weak notion of stability described above might be of considerable relevance. In fact, it is interesting to remark that such ideas are rather well common among applied physicists; indeed, people working in plasma physics and in the domain of molecular collisions are well aware of the fact that, for example in the collisions of identical diatomic molecules, the energy transfer between the translational degrees of freedom (center of mass energy) and the vibrational ones decreases exponentially with the vibrational frequency. In passing, one can make the technical remark that the method used by applied physicists to prove results of such a type is different from the one illustrated here, being based on some works of Landau [36] and Jeans [11,12]. Such a method (which we call the Jeans–Landau–Teller one) relies on the analyticity properties of the solutions of the differential equations of the problem, and makes essential use of the very well known property that the Fourier transform $\hat{f}(\omega)$ of an analytic function $f(t)$ decays exponentially with ω. In fact, the proofs usually found have a heuristic character; for a rigorous discussion see [37].

In any case, apart from this technical aspect, we would like to stress that the kind of results illustrated here, being based only on the consideration of simple general properties of the Hamiltonian dynamics, appear to have a rather general validity, and to be of interest for a large class of physical models. So, they should be relevant not only for applied physics, but rather for the general foundations of physics itself (see for example [38]).

References

[1] Nekhoroshev N.N., 1977, Usp. Mat. Nauk 32, [1977, Russ. Math. Surv. 32, 1]
[2] Nekhoroshev N.N., 1979, Trudy Sem. Petrows. No. 5, 5 [1985, Topics in Modern Mathematics, Petrovskii Sem. no. 5, O.A. Oleinik ed. (Consultant Bureau, New York)]
[3] Benettin G., Galgani L., Giorgilli A., 1985, Celestial Mechanics 37, 1
[4] Benettin G., Gallavotti G., 1985, J. Stat. Phys. 44, 293
[5] Benettin G., Galgani L., Giorgilli A., 1989, Comm. Math. Phys. 121, 557
[6] Benettin G., Galgani L., Giorgilli A., 1987, Comm. Math. Phys. 113, 87
[7] Moser J., 1955, Nachr. Akad. Wiss. Gottingen Math. Phys. Kl. 2a, 87
[8] Littlewood J.E., 1959, Proc London Math. Soc. 9, 343 and 525
[9] Boltzmann L., 1895, Nature 51, 413
[10] Boltzmann L., 1966, *Lectures on gas theory,* translated by S.G.Brush, University of Cal. Press; see especially section 45, *Comparison with experiments.*
[11] Jeans J.H., 1903, Phil. Mag. 6, 279
[12] Jeans J.H., 1905, Phil. Mag. 10, 91
[13] O'Neil T.M., Hjorth P.G., 1985, Phys. Fluids A 28
[14] O'Neil T.M., Hjorth P.G., Beck B., Fajans J., Malmberg J.H., 1990, *Collisional Relaxation of Strongly Magnetized Pure Electron Plasma (Theory and Experiment),* in Strongly coupled Plasma Physics, Proceedings of the Yamada Conference N. 24, Japan, North–Holland (Amsterdam), p. 313
[15] Rapp D., 1960, Journ. Chem. Phys. 32, 735
[16] Rapp D., Kassal T., 1969, Chem. Rev. 65, 61
[17] Benettin G., 1988, Nekhoroshev-like Results for Hamiltonian Dynamical Systems, lectures given at the Noto School "Non-Linear Evolution and Chaotic Phenomena", G. Gallavotti and A.M. Anile Editors (Plenum Press, New York)
[18] Giorgilli A., 1988, Relevance of Exponentially Large Time Scales in Practical Applications, Effective Fractal Dimension in Conservative Dynamical Systems, lectures given at the Noto School "Non-Linear Evolution and Chaotic Phenomena", G. Gallavotti and A.M. Anile Editors (Plenum Press, New York)
[19] Galgani L., 1988, *Relaxation Times and the Foundations of Classical Statistical Mechanics in the Light of Modern Perturbation Theory,* lectures given at the Noto School *'Non-Linear Evolution and Chaotic Phenomena,* G. Gallavotti and A.M. Anile Editors (Plenum Press, New York)
[20] Giorgilli A., 1988, *Rigorous results on the power expansions for the integrals of a Hamiltonian system near an elliptic equilibrium point,* Ann. Inst. H. Poincaré, 48 n. 4, 423–439
[21] Giorgilli A., Posilicano A., 1988, *Estimates for normal forms of differential equations near an equilibrium point,* ZAMP 39, 713–732
[22] Giorgilli A., Delshams A., Fontich E., Galgani L., Simò C., 1989, Effective Stability for a Hamiltonian System near an Elliptic Equilibrium Point, with an Application to the Restricted three Body Problem, J.Diff.Eq., 77, 167–198
[23] Simó C., 1989, Memorias de la Real Acad. Cienc. Art. Barcelona 48, 303
[24] Celletti A., Giorgilli A., 1991, On the stabililty of the Lagrangian points in the spatial restricted problem of three bodies, Cel. Mech. 50, 31–58
[25] Ulam S., 1965, preface to the Fermi, Pasta, Ulam paper, in E. Fermi, *Collected Papers* (Chicago).
[26] Fermi E., Pasta J., Ulam S., 1955, Los Alamos Report No. LA-1940, later published in, 1965, E. Fermi, *Collected Papers* (Chicago), and 1974, Lect. Appl. Math. 15, 143

[27] Benettin G., 1986, *Ordered and Chaotic Motions in Dynamical Systems with Many Degrees of Freedom,* in *Molecular-Dynamics simulation of Statistical Mechanical Systems*, Rendiconti della Scuola Italiana di Fisica "E. Fermi", G. Ciccotti and W.G. Hoover editors (North–Holland, Amsterdam)
[28] Izrailev F.M., Chirikov B.V., 1966, Sov. Phys. Dokl. 11, 30
[29] Bocchieri P., Scotti A., Bearzi B., Loinger A., 1970, Phys. Rev. A 2, 2013
[30] Cercignani C., Galgani L., Scotti A., 1972, Phys. Lett. A38, 403
[31] Galgani L., Scotti A., 1972, Phys. Rev. Lett. 28, 1173
[32] Benettin G., Galgani L., Giorgilli A., 1984, Nature 311, 444
[33] Benettin G., Galgani L., Giorgilli A., 1987, Phys. Lett. A 120, 23
[34] Galgani L., Giorgilli A., Martinoli A., Vanzini S., 1992, *On the problem of energy equipartition for large systems of the Fermi–Pasta–Ulam type: analytical and numerical estimates*, Physica D 59, 334–348
[35] Bambusi D., Giorgilli A., 1993, *Exponential stability of states close to resonance in infinite dimensional Hamiltonian Systems*, J. Stat. Phys. 71, 569
[36] Landau L.D., Teller E., 1936, Physik. Z. Sowjetunion 10, 34, and, 1965, in D. ter Haar ed. *Collected Papers of L.D. Landau*, Pergamon Press (Oxford), p. 147
[37] Carati A., Benettin G., Galgani L., 1992, *Towards a rigorous treatment of the Jeans–Landau–Teller method for the energy exchanges of harmonic oscillators*, Comm. Math. Phys. 150, 331–336
[38] Galgani L., 1992, *The quest for Planck's constant in classical physics*, in *Probabilistic methods in mathematical physics*, F. Guerra, M. Loffredo and C. Marchioro eds., World Scientific (Singapore)

Recent Developments in the Dynamics of Nonlinear Hamiltonian Systems with Many Degrees of Freedom

Marco PETTINI [1 2]

[1] Osservatorio Astrofisico di Arcetri, Largo Enrico Fermi 5, 50125 Firenze, Italy
[2] Istituto Nazionale di Fisica Nucleare, Sezione di Firenze, Italy

Abstract. A concise account is given here of some recent results concerning the dynamical properties of Hamiltonian systems with many degrees of freedom. Some of the main theoretical points of this research field are also briefly discussed and compared with the outcomes of recent numerical simulations.

New results, obtained with a differential geometrical approach to Hamiltonian dynamics, are also presented together with a mention of their consequences for a deeper understanding of the dynamical properties of self-gravitating N-body systems.

1. Introduction

Hamiltonian dynamics underlies almost all our descriptions of the physical world, both classical and quantum.

There are two opposite attitudes to cope with the description of the dynamics of generic Hamiltonian systems, let us call them, just for historical reasons, the Statistical Mechanics approach and the Celestial Mechanics approach. The former makes use of the "Stosszahlansatz" (molecular chaos hypothesis) – that becomes the "ergodic hypothesis" in ensemble theory – in order to circumvent the problem of knowing the dynamical details of a huge amount of particles. Whereas the latter has been developed to describe perturbed but regular motions of the celestial bodies belonging to our planetary system. Equilibrium Statistical Mechanics is a very powerful and well developed theory based on the assumption that the dynamics of generic Hamiltonian systems is ergodic, this means that *static* (or ensemble) averages can be used to predict the results of the experimental measurements of physical observables (an experimental apparatus performs a time average of a given observable). On the contrary Celestial Mechanics aims at working out the solutions of the equations of motion of the interacting bodies; this can be done only in an approximate way and the techniques developed to this purpose form the so called Classical Perturbation Theory (CPT).

However, there are cosmic systems made out of a very large number of gravitationally interacting bodies, such are globular clusters and galaxies, so that CPT may fail also in describing celestial systems.

Typical time scales of condensed matter systems, those commonly described by equilibrium Statistical Mechanics, may range from picoseconds down to femtoseconds so that equilibrium generally sets up in very short times compared with our macroscopic point of view. At variance, for the above mentioned stellar systems the concept of equilibrium becomes more fuzzy and questionable because of very large characteristic time scales, in this case neither CPT nor equilibrium Statistical Mechanics may be adequate theoretical tools: here the central role is certainly played by the *dynamics.*

By means of the word *equilibrium* we denote a state that can be fully described by a one-parameter *sufficient statistics*, the temperature, in the sense of mathematical statistics[1].

Studying the dynamics of gravitationally interacting N-body systems is certainly a very difficult and challenging task. To begin with, there is nothing like the Debye screening, the gravitational potential is singular and even the numerical integration of the equations of motion is rather delicate.

After Miller's pioneering computer simulations[2] of self-gravitating systems, a *conceptual laboratory* became available to the theoretical investigation.

In order to get a meaningful insight of the physics underlying any observable phenomenology, the theoretical models must try to get rid of a good deal of details of a real system by seeking some hierarchy in the physical processes involved. Simplifying assumptions are therefore essential. In fact we do not need a computer simulation of many billions of gravitationally interacting stars to *understand* the dynamic properties or the morphogenetic mechanisms of a galaxy: this would be not only practically impossible but even unintelligent.

On the other hand, theoretical progress can benefit very much of the use of those "toy models" that are designed to shine some light on difficult problems of unknown solutions and hard to grasp even heuristically.

In such a context it is worth reminding that since the dawning of the computer era, scientists like J. von Neumann, S. Ulam and E. Fermi realized how far reaching the potentialities were of the new instrument.

With the following illuminating words S. Ulam remembered how the collaboration with E. Fermi began:

"After the war, during one of his frequent summer visits to Los Alamos, Fermi became interested in the development and potentialities of the electronic computing machines. He held many discussions with me on the kind of future problems which could be studied through the use of such machines. We decided to try a selection of problems for heuristic work where in the absence of closed analytic solutions experimental work on a computing machine would perhaps contribute to the understanding of properties of solutions. This could

be particularly fruitful for problems involving the asymptotic – long time or "in the large" – behaviour of nonlinear physical systems. In addition, such experiments on computing machines would have at least the virtue of having the postulates clearly stated. This is not always the case in an actual physical object or model where all the assumptions are not perhaps explicitly recognized.

Fermi expressed often the belief that future fundamental theories in physics may involve nonlinear operators and equations, and that it would be useful to attempt practice in the mathematics needed for the understanding of nonlinear systems. The plan was then to start with the possibly simplest such physical model and to study the results of the calculation of its long-time behaviour. Then one would gradually increase the generality and the complexity of the problem calculated on the machine... perhaps problems of pure kinematics, e.g., the motion of a chain of points subject only to constraints but no external forces, moving on a smooth plane convoluting and knotting itself indefinitely. These were to be studied preliminary to setting up ultimate models for motion of systems where "mixing" and "turbulence" would be observed. The motivation then was to observe the rates of mixing and "thermalization" with the hope that the calculational results would provide hints for a future theory. One could venture a guess that one motive in the selection of problems could be traced to Fermi's early interest in the ergodic theory..."[3].

Actually, a famous numerical work by Fermi, Pasta & Ulam[3] on a chain of oscillators coupled by a weak anharmonicity, since then called FPU model, represents a milestone for nonlinear dynamics and particularly for nonlinear Hamiltonian systems with many degrees of freedom.

These considerations aim at warning about the need for a tight interplay between all the existing analytic tools to tackle Hamiltonian dynamics and numerical simulations. Such an interplay has been very fruitful in the case of condensed matter systems and is discussed in the following. At a first sight the subject might seem very far from the gravitational N-body problem, however this is a glance in a "theoretical tool-box" that has been recently filled with new tools; in the final part of the present contribution it is shown that these new tools are definitely interesting for the gravitational N-body problem.

2. Outline of Analytical and Numerical Results

As already reminded above, historically, Hamiltonian dynamics has received its strongest impulse from Celestial Mechanics. As the motions of celestial bodies are mainly regular, at least on our observational time scales, it is readily understood why CPT deals with quasi-integrable systems, i.e., described by Hamiltonians that in action-angle coordinates read as

$$H(\theta, \mathbf{I}) = H_0(\mathbf{I}) + H_1(\theta, \mathbf{I}) , \quad \sigma = \frac{||H_1||}{||H_0||} \ll 1 . \tag{1}$$

Within the framework of CPT, the motions of Hamiltonian (1) are obtained by seeking the generating function of a canonical transformation $S : (\theta, \mathbf{I}) \rightarrow (\alpha, \mathbf{J})$ that brings H to

$$H'(\alpha, \mathbf{J}) = H'_0(\mathbf{J}) + H'_1(\alpha, \mathbf{J}) \tag{2}$$

with $||H'_1||/||H'_0|| = O(\sigma^2)$, so that the motions of $H'_0(\mathbf{J})$ can approximate the motions of $H(\theta, \mathbf{I})$ up to times of the order of $1/\sigma^2$.

This procedure can be iterated to higher orders; at each step the motion can be approximately integrated up to longer and longer times as $1/\sigma^2, \ldots 1/\sigma^n \ldots$. Serious difficulties for CPT are represented by the appearance of small denominators (resonances) in the Fourier development of S; these small denominators are responsible for the divergence of the perturbative series.

Kolmogorov[4] found a superconvergent technique which can overcome such divergent terms, provided that some condition is fulfilled by the resonant moduli. The development of Kolmogorov stability theorem by Arnold[5] and Moser[6] led to the so called KAM theorem (or theory), which states that sufficiently irrational tori can survive –being only deformed– to a non-integrable perturbation $H_1(\theta, \mathbf{I})$. Moreover, for $0 \leq \sigma < \sigma_c$, the set of stable invariant tori is an open set, i.e., of positive Lebesgue measure. This result holds true in spite of Poincaré-Fermi non existence theorem[7], which states that at $N \geq 3$ generic non integrable Hamiltonians, like in Eq. (1), cannot have any analytic integral of motion besides energy. Poincaré-Fermi theorem is invoked to corroborate the widespread point of view that generic perturbed systems are always ergodic, and that the so called ergodic problem at the grounds of statistical mechanics does not represent a serious problem –at least for physics. Doubts about this point of view could be raised by KAM theorem, but the strong N-dependence of the KAM threshold σ_c, below which absolute stability of an open set of tori can subsist, prevents any difficulty for ergodicity already at few tenths of degrees of freedom. Typically it is found $\sigma_c(N) = A \exp(-N^2 \log N)$ or also[8] $\sigma_c(N) = B \exp(-N \log N)$. For this reason, it is clear that KAM theorem, even if it represents a real breakthrough from a theoretical standpoint, is of no practical use for large systems.

A new remarkable approach to the stability problem is represented by Nekhoroshev theorem[9]. In this case, instead of looking for infinite time stability of a set of invariant tori, one simply tries to estimate the stability time scale of a given initial condition. Denoting by $\mathbf{I}(0)$ such an initial condition, Nekhoroshev theorem states that

$$|\mathbf{I}(t) - \mathbf{I}(0)| < M\sigma^a \tag{3}$$

holds true at least up to a time $t = \tau_0 \exp(\sigma_0/\sigma)^\gamma$, with τ_0 and σ_0 suitable constants. This gives a lower bound for the diffusion time of the actions when a perturbation is present.

Nekhoroshev theorem has had a great heuristic and conceptual impact, but it has a major difficulty originated by the strong N-dependence of the exponent γ in the exponential bound for the stability time. In the original derivation of the theorem, it was found $\gamma(N) \sim 1/N^2$. Subsequent efforts[10] to improve this estimate always gave $\gamma \sim 1/N$. This last estimate is considered optimal for generic systems and confirmed optimal by numerical tests designed to make a direct verification of the analytical results. Therefore, also Nekhoroshev theorem is not very useful for large N systems.

Finally, the main limitation in applying perturbation theory to generic large N systems is conceptual, in fact CPT is conceived to deal with the regular (or almost regular) regions of phase space. But it is a fact that the resonant manifolds of $H_0(\mathbf{I})$, defined by $\sum_j k_j(\partial H_0/\partial I_j) = 0$ (k_j are integers), have a dense and connected set of intersections with the constant energy surface. This set of resonances is the backbone of a stochastic web which is generated by the action of $H_1(\theta, \mathbf{I})$.

In general case, at $N \gg 3$, the presence of this connected stochastic web on the energy surface, the absence of analytic or smooth integrals besides energy (according to Poincaré-Fermi theorem) and the extreme smallness of the KAM threshold, justify the statement that generic nonintegrable Hamiltonian systems are ergodic, at least in the common physical sense. Therefore if we adopt the point of view of perturbation theory, the dynamical properties of high dimensional systems appear rather trivial. On the contrary, if the stochastic web undergoes some structural change by varying the energy density of the system, the corresponding change of the topology of phase paths would yield different diffusion regimes in phase space. Such a possibility has interesting physical consequences. For example, a fast diffusion process, due to *strong chaos*, implies fast mixing and can be related to a *strong* overlapping of resonances in the stochastic web. In this case diffusion takes place in every direction on the energy surface. At variance, a *weak* overlapping of resonances within the stochastic web, or *weak chaos*, makes dominant the diffusion *along* resonances. In this case the phase space paths get more and more tortuous, thus the mixing rate is affected and becomes extremely long at low energy.

Let us illustrate in few more details how such a phenomenology actually shows up, what kind of physics is thus involved, and how a Riemannian description of Hamiltonian chaos – though still in infancy – provides an explanation of numerical simulations.

In the following we deal with the FPU β-model only. This model is considered a paradigmatic one because it shares with many other systems (Morse and Lennard-Jones chains, lattice φ^4 model, coupled rotators) the phenomenology reported below, and in addition it has the peculiar prop-

erty that several exact analytic computations can be performed only for it. Needless to say how important this is in order to increase the confidence in numerical simulation results.

The FPU β-model is described by the Hamiltonian

$$H = \sum_{i=1}^{N} \left[\frac{1}{2} p_i^2 + \frac{1}{2}(q_{i+1} - q_1)^2 + \frac{\mu}{4}(q_{i+1} - q_1)^4 \right] \tag{4}$$

which, using the transformation $q_i = (2/N)^{1/2} \sum_{k=1}^{N} Q_k \sin(ik\pi/N)$, can be cast in the form

$$H(\mathbf{P}, \mathbf{Q}) = \sum_{k=1}^{N} \left[\tfrac{1}{2}(P_k^2 + \omega_k^2 Q_k^2) + \mu \sum_{k_1,k_2,k_3=1}^{N} C(k, k_1, k_2, k_3)\, Q_k Q_{k_1} Q_{k_2} Q_{k_3} \right] \tag{5}$$

where the coefficients (given for simplicity in the case of fixed boundary conditions) are $C(k, k_1, k_2, k_3) = \frac{1}{8N} \omega_k \omega_{k_1} \omega_{k_2} \omega_{k_3} \delta(k + k_1 + k_2 + k_3)$ and $\omega_k = 2\sin(\pi k/N)$.

Equation (5) shows that the FPU model, rewritten in normal mode coordinates, represents a collection of harmonic oscillators coupled by a simple anharmonicity familiar in condensed matter physics. Apart from the particular form of the coefficients $C(k, k_1, k_2, k_3)$, typical of the FPU model, this is the functional form of a generic Hamiltonian that describes a collection of interacting quasi-particles (phonons, excitons, polarons, magnons etc.).

For practical reasons it is much more convenient to integrate the equations of motion that are derived from Hamiltonian (4). The numerical integration algorithm *must* be symplectic. A symplectic algorithm performs at each time step a canonical transformation of variables: $(\mathbf{p}(t), \mathbf{q}(t)) \to (\mathbf{p}(t+\Delta t), \mathbf{q}(t+\Delta t))$ thus providing a faithful representation of an Hamiltonian flow. The well known leap-frog algorithm is certainly a simple but excellent example of an explicit symplectic algorithm that can be easily extended at any order and improved in the so-called "bilateral" version. Details about this point can be found in [11].

Integrating the equations of motion of the FPU model, or of the other mentioned models, does not make any serious problem because this kind of systems represents a set of particles that interact through a nearest-neighbour potential with minimum (this is the main difference with the gravitational N-body problem).

The numerical experiment performed by Fermi, Pasta and Ulam produced a striking result: the energy, initially given to one of the normal modes of a chain of 64 particles, was exchanged in a complicated but apparently recurrent way among all the other modes; adversely to the expectations, no tendency toward equipartition of energy was observed. Thereafter people spoke of the "FPU problem", and two different approaches were adopted in the sake of an explanation of the lack of equipartition. The first attempt concerns the integrability of nonlinear PDEs. In fact a good approximation of the FPU

model in the continuum limit is given by a modified version of the Korteweg-de Vries equation where Zabusky & Kruskal[12] discovered the existence of solitons. The second approach tackles the problem from the opposite side, that of stochasticity[13]. Here it was suggested that some kind of threshold phenomenon might exist according to the wavelength of the initially excited mode: a critical value of the excitation amplitude below which the motion is ordered. The idea of a stochastic transition seemed in agreement with several numerical experiments[14] though, strictly speaking, it is not correct.

Even if Kolmogorov's theorem was contemporary to the FPU experiment, the two works were related for the first time by Izrailev & Chirikov several years later[13].

However KAM theorem cannot explain the existence of any threshold observed at rather large perturbation amplitudes in numerical experiments because – as it became clear much later – σ_c drops down to values which are exceedingly small at large N. Just to give an idea of how small the threshold becomes, already at $N = 100$ we get the rough estimate $\sigma_c \sim e^{-400}$!

Subsequent numerical simulations, thanks to the increasing power of computers, pointed out better and better that at low energy – even with large N – *apparently* regular motions subsist, while at high energy – above some threshold – the motions become fully chaotic and quick mixing takes place. Many authors[15] reported at worst only weak tendencies, if any, of this threshold to shrink to zero by increasing N.

More recent and extensive numerical experiments[16,17], confirm, as expected, that the dynamics is always chaotic and that we can reasonably consider ergodic, in the physical sense, large N systems like the FPU lattice. The property that drastically changes when the specific energy exceeds some threshold value is rather the *mixing rate*, which in turn is related with the kind of chaoticity of the system (weak or strong). In other words, by repeating with nowadays computers the original FPU simulations, one finds that after a sufficiently long time equipartition of energy is always attained. The lack of equipartition in the FPU original experiment was only due to the choice of the energy value in the slow mixing regime. What is nowadays evident[16,17] – for nonlinear Hamiltonian systems with $N \geq 3$ – is that a Strong Stochasticity Threshold (SST) exists between these two regimes of weak and strong chaos.

The discovery of this threshold (SST) has clarified the origin of an ubiquitous bimodality in the dynamical behaviour of nonlinear, nonintegrable Hamiltonian systems. Such a bimodality, observed in numerical simulations by varying the energy, appeared – at different epochs – as an effect of the mentioned stochasticity threshold, or of some equipartition threshold[18], or due to the slowing down of phase space diffusion as described by Nekhoroshev theorem[9]. However, because of the above mentioned reasons, the only possibility which is left open to explain such a bimodality is to account for the existence of qualitatively different regimes of chaoticity.

A good definition of the SST could be given through the ϵ-behaviour of any observable that is sensitive to the difference between weak and strong chaos (ϵ is the energy per degree of freedom).

However, the best definition is provided by the crossover value ϵ_c of the scaling of the largest Lyapunov exponent, $\lambda_1(\epsilon)$. In fact, this is an unambiguous definition, directly related with the level of chaoticity, and is *independent of the choice of initial conditions*[17] (also random initial conditions at equipartition are chosen). Because of this last property, the crossover of $\lambda_1(\epsilon)$ gives an *intrinsic* definition of a *global* transitional feature of the dynamics, and this is why $\lambda_1(\epsilon)$ is markedly superior to any other observable.

Strong chaoticity in the FPU model is quantitatively described by the scaling law $\lambda_1(\epsilon) \sim \epsilon^{2/3}$ of the largest Lyapunov exponent. Whereas weak chaoticity corresponds to $\lambda_1(\epsilon) \sim \epsilon^2$ at $\epsilon < \epsilon_c$. The $\epsilon^{2/3}$-law is explained by a random matrix approximation of the tangent dynamics as follows. Let

$$\mathbf{M}(\Omega) = \begin{pmatrix} \mathbf{I} & \tau\mathbf{I} \\ \tau\Omega & \mathbb{I} + \tau^2\Omega \end{pmatrix} \tag{6}$$

be the Jacobian of the discretized Hamiltonian flow derived from Eq. (4) with $\Omega_{ij} = -\partial^2 U(\mathbf{q})/\partial q_i \partial q_j$ and τ a discretization time (for instance the time integration step). $\mathbf{M}$ is a $2N \times 2N$ symplectic matrix which maps a vector $\boldsymbol{\xi}(t)$ tangent to the flow into a vector $\boldsymbol{\xi}(t+\tau)$. The largest Lyapunov exponent λ_1 is given by

$$\lambda_1 = \lim_{n\to\infty} \frac{1}{n\tau} \ln \left\langle \left\{ \frac{\boldsymbol{\xi}^T(0) \left[\prod_{j=1}^{n} \mathbf{M}^T(\mathbf{q}(j\tau))\right] \cdot \left[\prod_{k=1}^{n} \mathbf{M}(\mathbf{q}(k\tau))\right] \boldsymbol{\xi}(0)}{\boldsymbol{\xi}^T(0) \cdot \boldsymbol{\xi}(0)} \right\}^{\frac{1}{2}} \right\rangle . \tag{7}$$

The matrix elements of Ω contain terms like $(q_{i+1} - q_i)^2$. In the random matrix approximation the hypothesis of δ-correlation in time is made for the fluctuating part $\tilde{\Omega}$ of Ω, i.e., $\langle \tilde{\Omega}_{ij}(k\tau)\, \tilde{\Omega}_{ij}(l\tau) \rangle = (\gamma_{ij}/\tau)\, \delta_{kl}$. The average $\langle \cdot \rangle$ in Eq. (7) is considered over different realizations of the random matrix process. The average of γ_{ij} is given by

$$\gamma = \frac{1}{N} \sum_{i=1}^{N} \left[(q_{i\pm1} - q_i)^2 - \langle (q_{i\pm1} - q_i)^2 \rangle \right]^2 . \tag{8}$$

The computation of Eq. (7) yields[16] $\lambda_1 \sim \gamma^{1/3}$. Then the average of $\gamma(\epsilon)$ can be computed on the constant energy surface[17] and finally yields $\langle \gamma \rangle(\epsilon) \sim \epsilon^2$, hence $\lambda_1 \sim \gamma^{1/3}(\epsilon) \sim \epsilon^{2/3}$. Therefore the meaning of $\lambda_1 \sim \varepsilon^{2/3}$ is that the corresponding dynamical behaviour is strongly chaotic and can be modeled by an uncorrelated random walk in phase space, which also means that diffusion

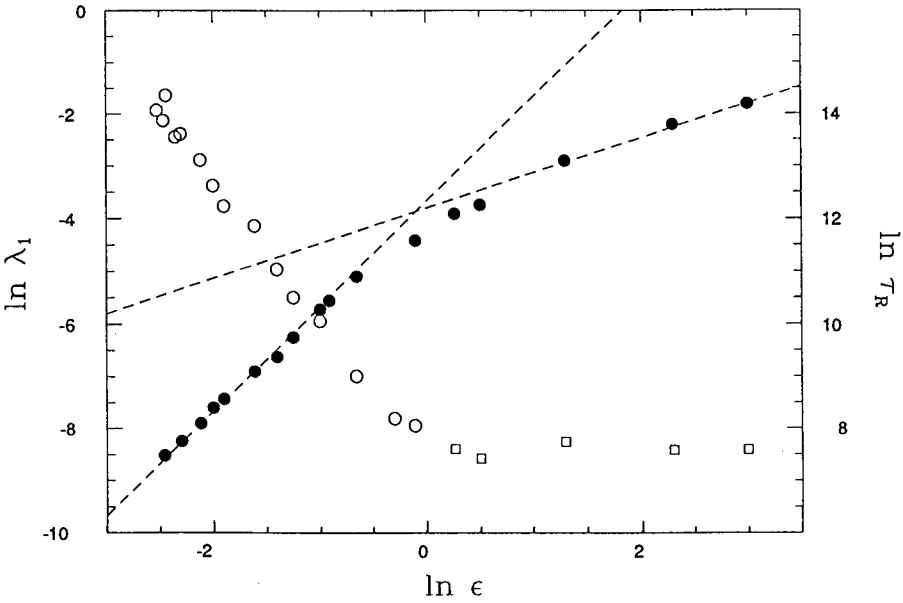


Fig. 1. Synopsis of the largest Lyapunov exponent λ_1 and relaxation time τ_R vs. energy density ε for the FPU model defined in Eq. (4) with $N = 128$. λ_1 is represented by full circles. Dashed lines are references to power laws ε^2 and $\varepsilon^{2/3}$. Open circles and squares represent relaxation times to equipartition of energy among normal modes of the energy initially concentrated in a wave-packet made with the four lowest modes of the chain. The crossover point of $\lambda_1(\varepsilon)$ defines the *strong stochasticity threshold* ε_c. Below ε_c the mixing time τ_R rapidly grows by decreasing ε

takes place in any direction on the energy surface. This corresponds to fast mixing for the overwhelming majority of initial conditions.

At $\epsilon < \epsilon_c$, it is evident that the random matrix approximation breaks down, phase space diffusion is much slower, chaoticity is definitely weaker and characterized by a steeper (and model dependent) scaling law $\lambda_1(\epsilon)$. In this regime, depending on the initial conditions, even very long mixing times can be observed. Phase space paths are now more tortuous and can even look regular when followed during an insufficiently long observational time.

The SST has some interesting consequences on the nonequilibrium behaviour of these systems: the energy dependence of the relaxation time, needed to reach equipartition from an initial state far from it, is very sensitive to the chaoticity regime. Let us look at Fig. 1 where a synopsis is given of the ϵ-dependence of the relaxation time τ_R and of the largest Lyapunov exponent λ_1 (in this case the four lowest modes were initially excited). At high energy, in the domain of strong chaos, we observe $\lambda_1(\epsilon) \sim \epsilon^{2/3}$ and τ_R is almost independent of the energy. At low energy, in the domain of weak chaos, $\lambda_1(\epsilon) \sim \epsilon^2$ i.e., at any $\epsilon < \epsilon_c$ the actual value of λ_1 is smaller than that obtained by extrapolating at low energy the $\epsilon^{2/3}$ behaviour; now τ_R is

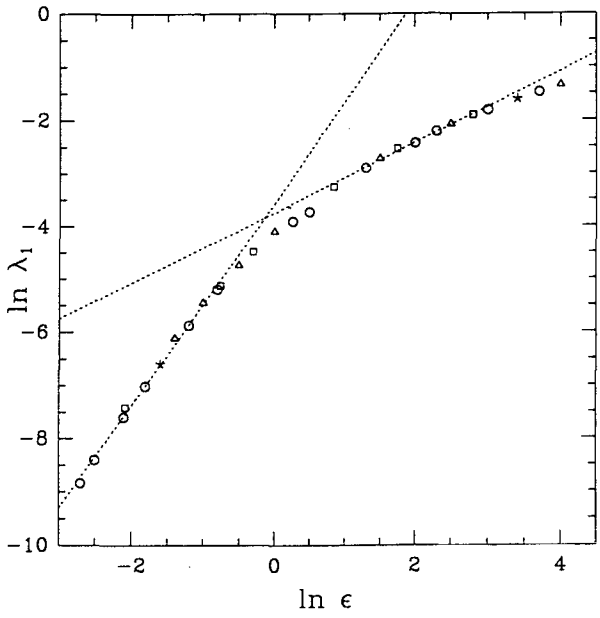


Fig. 2. FPU model. Largest Lyapunov exponent λ_1 versus energy density ϵ. Initial conditions: random at equipartition (circles); $\mathbf{n}_{\mathrm{exc}} = (30, 31, 32, 33)$ (squares); $\mathbf{n}_{\mathrm{exc}} = (54, 55, 56, 57)$ (triangles); $\mathbf{n}_{\mathrm{exc}} = (3, 38, 61)$ (asterisks). Dashed lines are references to power laws $\epsilon^{2/3}$ and ϵ^2. Independently of the initial condition, the crossover in the scaling behaviour $\lambda_1(\epsilon)$ always occurs at the same value of the energy density. This shows the *intrinsic* character of the SST. $N = 128$

strongly dependent on the energy and steeply increases by lowering ϵ. Notice that there is an evident relationship between τ_{R} and λ_1, but this is neither simple nor very direct. In other words, it is *not true* that λ_1^{-1} is τ_{R}, neither a simple functional relationship can be obtained, say, by dimensional arguments; the reason is that by changing the initial conditions $\tau_{\mathrm{R}}(\epsilon)$ changes whereas $\lambda_1(\epsilon)$ does not (see Fig. 2). Figure 3 shows how complicate the relationship is: τ_{R} is now reported at different energies as a function of the average wavenumber of the initially excited packet. It is now evident that above the SST the initial excitation of high frequency modes results in a faster equipartition, while below the SST the situation is reversed[17]. It is worth mentioning that, at any energy, the virialization time is extremely fast in these models, and that the relaxation time is also strongly dependent on the observable[18].

This point can be of interest for the gravitational N-body problem: it warns about possible oversimplifications in the definition of "the" relaxation time and of its relation with λ_1.

It is worth mentioning that a similar crossover in the scaling of λ_1 has been found in Morse and Lennard-Jones lattices[20], one-dimensional coupled rotators, and in a two-dimensional system of coupled rotators[21]; this last system is the classical counterpart of a 2D Heisenberg XY model. In correspondence with the critical temperature – at which the system undergoes a Kosterlitz-Thouless phase transition – also a transition appears between weak and strong chaotic behaviour. This suggests an intriguing relationship with

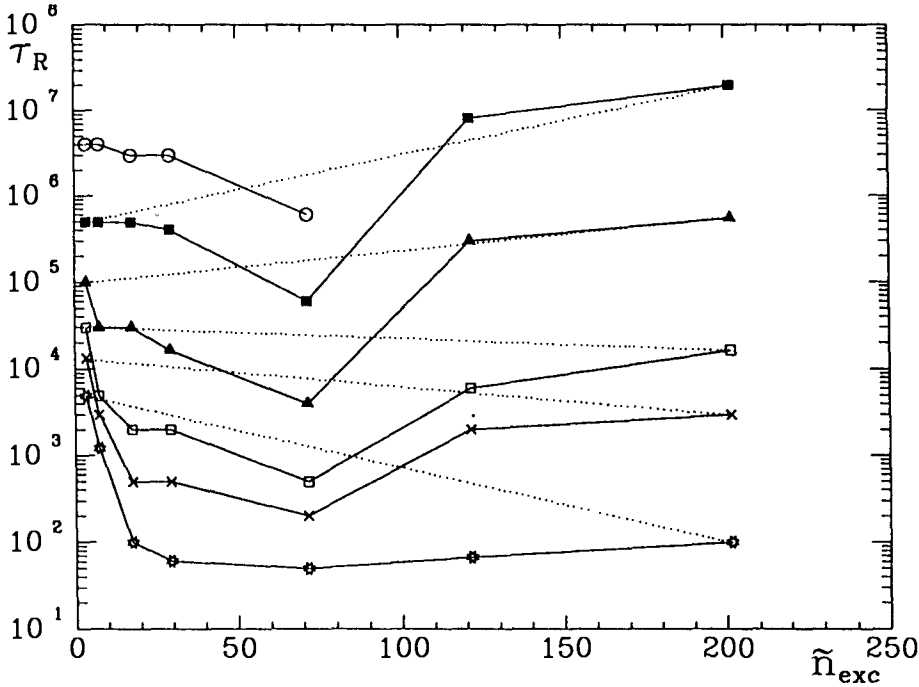


Fig. 3. FPU model. Equipartition time τ_R versus mean wavenumber $\tilde{n}_{exc}$ of the initially excited packet for different values of energy density ε. From top to bottom: $\varepsilon = 0.059, 0.137, 0.39, 1.66, 3.9, 54.7$. Each wavepacket is made out of four modes. Dotted lines join the lowest with the highest excited modes. These lines change the sign of their slopes around the SST. This fact puts in evidence that below the SST high frequency excitations are harder to relax than low frequency ones. The converse happens at $\varepsilon > \varepsilon_c$

the dynamical counterpart of phase transitions in systems with continuous symmetries.

Finally, there are also several experimental phenomena that suggest the existence of measurable physical effects of the SST. Let us just quote the sudden increase of the width of Raman spectral lines in molecular crystals; this occurs at some "critical" temperature below which the lines are narrow and the linewidth is rather insensitive to the temperature, and above which the lines become broader and the linewidth quickly increases with the temperature[22]. This phenomenon seems related with a different rate of chaoticity in the phonon dynamics thus with a different degree of energy exchange among the phonons, a fact that gives, as observable counterpart, different line broadenings.

3. Riemannian Description of Hamiltonian Chaos

Now, the problem that naturally arises is how to explain the origin of the Strong Stochasticity Threshold and how to compute, say, the crossover energy. According to the above given arguments, one has to look for some *non-perturbative* method. The only theoretical framework dealing with the opposite situation of CPT, that is with completely chaotic trajectories, is ergodic theory. A relevant contribution to this field was given by Krylov[23]. He realized that mixing, rather than ergodicity, is the most important property for statistical physics (i.e., convergence of time averages to ensemble ones in a finite time) and that the study of the mixing properties of a physical system could have taken advantage of already existing results about the stability properties of geodesics on Riemannian manifolds of negative curvature. These results, associated with the names of Hopf, Hadamard, Hedlund, had never been exploited before in physics.

The follow up of Krylov's ideas took place within ergodic theory, with the fundamental works by Sinai, Anosov and others on geodesic flows[24]. But generic Hamiltonian flows, for instance like that of the FPU model, have not been touched by these methods.

Taking advantage of the well known fact that the equations of motion of a mechanical system, according to Maupertuis' principle, are given by the extremals of the action integral

$$A = \int_{t_0}^{t_1} 2T(\mathbf{q}, \dot{\mathbf{q}})\, \mathrm{d}t \tag{9}$$

where T is the kinetic energy, the trajectories of a Newtonian system can be viewed as geodesics of a Riemannian manifold. This is possible because also for the geodesics a variational formulation can be given (stationarity of the arc-length functional).

Having set $E > V(\mathbf{q})$ for $\mathbf{q} \in M$, E is the energy, V is the potential and M is configuration space, the kinetic energy metric (or Jacobi metric, g_J) is given by

$$g_{ij} = \big(E - V(\mathbf{q})\big)\, a_{ij}\ , \tag{10}$$

where a_{ij} is out of $2T = a_{ij}\dot{q}^i\dot{q}^j$.

If we denote by Γ^i_{jk} the connection coefficients derived from the metric (9), the corresponding geodesics are given by

$$\frac{\mathrm{d}^2 q^i}{\mathrm{d}s^2} + \Gamma^i_{jk}(\mathbf{q})\frac{\mathrm{d}q^j}{\mathrm{d}s}\frac{\mathrm{d}q^k}{\mathrm{d}s} = 0 \tag{11}$$

and one easily recovers

$$\frac{\mathrm{d}^2 q^i}{\mathrm{d}t^2} = -\frac{\partial V}{\partial q^i} \tag{12}$$

i.e., Newton's equations of motion.

As already pointed out in [25,26], there are several possibilities of rephrasing Hamiltonian dynamics in geometric terms by choosing as ambient manifold: *i)* configuration space M equipped with the Jacobi metric g_J; *ii)* configuration space time $M \times \mathbf{R}$ with the structure of a Finsler space induced by a suitable metric g_F; *iii)* enlarged configuration space-time $M \times \mathbf{R} \times \mathbf{R}$ equipped with Eisenhart metric; *iv)* the tangent bundle TM of configuration space equipped with the Sasaki lift g_S of g_J and suitable restrictions to TM_E.

Let us just mention that to derive a Riemannian structure from Hamilton least action principle, after Eisenhart theorem we must consider an enlarged configuration space-time $M \times \mathbf{R}^2$ whose local coordinates are $q^0, q^1, \ldots, q^i, \ldots, q^N, q^{N+1}$, with $(q^1, \ldots, q^N) \in M$, $q^0 \in \mathbf{R}$ is the time coordinate and $q^{N+1} \in \mathbf{R}$ is given by

$$q^{N+1}(t) = C_1^2 t + C_2 - \int_0^t L(\mathbf{q}, \dot{\mathbf{q}})\, \mathrm{d}t \,, \tag{13}$$

where C_1 and C_2 are arbitrary constants. For standard Hamiltonian functions $H = T + V(\mathbf{q})$, in this coordinate system the arc-length of the manifold $M \times \mathbf{R}^2$ is given by

$$\mathrm{d}s_E^2 = a_{ij}\mathrm{d}q^i\mathrm{d}q^j - 2V(\mathbf{q})(\mathrm{d}q^0)^2 + 2\mathrm{d}q^0\mathrm{d}q^{N+1} \,, \tag{14}$$

where a_{ij} is the kinetic energy matrix. Again the geodesics of this manifold are the natural motions of the given Hamiltonian system. Among the remarkable properties of Eisenhart metric, we mention that it makes possible to give a new meaning to the widespread numerical algorithm to compute Lyapunov exponents; details can be found in [25].

The stability properties of geodesics can be studied by means of the Jacobi–Levi-Civita equation for the second variation of the action integral. This equation links the divergence of nearby geodesics to the curvature properties of the underlying manifold through the Riemann curvature tensor R^i_{jlk} and reads as

$$\frac{\nabla}{\mathrm{d}s}\frac{\nabla}{\mathrm{d}s}\xi^i + R^i_{jlk}(\mathbf{q})\frac{\mathrm{d}q^j}{\mathrm{d}s}\xi^l\frac{\mathrm{d}q^k}{\mathrm{d}s} = 0 \,, \tag{15}$$

where ξ is the vector field of geodesic separation and can be used to measure the distance between nearby geodesics; $(\nabla\xi/\mathrm{d}s)$ is the covariant derivative along a geodesic.

Now the question is: what is the relationship between instability of the geodesics and chaos?

First of all remember that deterministic chaos is originated whenever the following two basic (topological) ingredients are somehow realised: *stretching* and *folding* of a given set of initial conditions; these are necessary for the appearance of a hyperbolic limit set[27].

The traditional explanation of the origin of chaos in Hamiltonian systems is based upon the picture of homoclinic intersections.

Homoclinic intersections, though of little practical (computational) use at high dimension, are responsible for the appearance of a hyperbolic invariant set (Smale-Birkhoff theorem) deeply affecting the dynamics. Therefore Hamiltonian chaos too stems from the two mentioned ingredients – *stretching* and *folding* of volumes in phase space.

In the Riemannian description of Hamiltonian chaos, *stretching* is provided by the *instability* of nearby geodesics, and *folding* by not allowing the distance between them to grow indefinitely, that is by compactedness of the ambient manifold. Under these circumstances, the phase trajectories are compelled to fold themselves in a very complicated manner so that their behaviour becomes practically unpredictable.

This approach to the explanation of the origin of Hamiltonian chaos is alternative to the standard approach of homoclinic intersections. The geometric description has the following great advantages: *i)* it is a nonperturbative approach, i.e., it holds true at any energy and at any degree of nonlinearity; *ii)* it makes use of the natural coordinates of a system, there is no need for action-angle coordinates; *iii)* it unifies the explanation of the origin of chaos with the method of measuring its strenght.

For the majority of systems of physical interest the configuration space manifold is compact, that is the coordinates remain bounded during their time evolution; in principle this is not the case of the self-gravitating N-body systems, however, if the characteristic time of mass evaporation out of the system is much longer than the dynamical instability time, then we can still consider that instability originates chaos.

The instability (stretching) of nearby geodesics is described by Eq. (15), whence in principle N "geometric" instability exponents could be obtained by averaging – along a trajectory – the eigenvalues of the matrix $Q^i_j = R^i_{ljk}\dot{q}^l\dot{q}^k$. However, due to the large number of independent components of the Riemann tensor, a practical difficulty arises. Therefore one is led to seek a more handy version of Eq. (15).

The first and most natural idea is to replace the system of N evolution equations for $\xi^1 \ldots \xi^N$ – given by (15) – with a single scalar equation describing the evolution of the norm of the separation vector $\|\xi\|^2$. In this way some information will be lost but knowing the evolution of the distance between two nearby geodesics is sufficient to describe and measure their degree of instability.

Standard algebraic manipulations[28] of Eq. (15) yield the exact result

$$\frac{1}{2}\frac{\mathrm{d}^2\|\xi\|^2}{\mathrm{d}s^2} + \left(R_{ijkl}\,\mu^i\frac{\mathrm{d}q^j}{\mathrm{d}s}\mu^k\frac{\mathrm{d}q^l}{\mathrm{d}s}\right)\|\xi\|^2 - \left\|\frac{\nabla\xi}{\mathrm{d}s}\right\|^2 = 0 \tag{16}$$

where $\mu^i = \xi^i/\|\xi\|$ are the components of the unit vector codirectional with ξ. Unless we can rewrite this equation in closed form it is useless.

In order to work out a scalar equation for $\|\xi\|^2$ in closed form, we must operate two approximations. First, we use the inequality

$$\left\| \frac{\nabla \xi}{\mathrm{d}s} \right\|^2 \geq \left(\frac{\mathrm{d}}{\mathrm{d}s} \|\xi\| \right)^2 \tag{17}$$

to replace the last term in Eq. (16). Second, we need to replace $R_{ijkl}\mu^i \dot{q}^j \mu^k \dot{q}^l$ with something independent of μ (or, equivalently, ξ). To this purpose, we recognize that this quantity is just the sectional curvature $K^{(2)}$ associated with the two dimensional plane spanned by ξ and $\mathrm{d}\mathbf{q}/\mathrm{d}s$. Of course, such a curvature depends upon the particular choice of ξ and $\mathrm{d}\mathbf{q}/\mathrm{d}s$ at a given point of M and still requires the knowledge of the evolution of ξ^i (or, equivalently, μ^i). Thus, at any point of M, we need to replace $K^{(2)}(\xi, \mathrm{d}\mathbf{q}/\mathrm{d}s)$ by some local average. Take, for simplicity, $\xi \perp \mathrm{d}\mathbf{q}/\mathrm{d}s$. There are at hand two possibilities. The first choice consists in averaging – with some uniform distribution – all the sectional curvatures given by the $N(N-1)$ mutual orientations of ξ and $\mathrm{d}\mathbf{q}/\mathrm{d}s$ at any point, and then replacing $K^{(2)}(\xi, \mathrm{d}\mathbf{q}/\mathrm{d}s)$ by such a local mean. Remind that, if we denote by $\xi_{(a)}$ and $(\mathrm{d}\mathbf{q}/\mathrm{d}s)_{(b)}$, $a, b = 1, \ldots, N$, the N independent choices of each vector, then it is

$$K^{(2)}\left(\xi, \frac{\mathrm{d}\mathbf{q}}{\mathrm{d}s}\right) \|\xi\|^2 \left\| \frac{\mathrm{d}\mathbf{q}}{\mathrm{d}s} \right\|^2 = R_{ijkl}\xi^i_{(a)} \left(\frac{\mathrm{d}q^j}{\mathrm{d}s}\right)_{(b)} \xi^k_{(a)} \left(\frac{\mathrm{d}q^l}{\mathrm{d}s}\right)_{(b)} , \tag{18}$$

and it is well known that

$$\sum_{a,b}^{1\ldots N} K^{(2)}_{ab} = \mathcal{R} . \tag{19}$$

This means that $\mathcal{R}/N(N-1)$ at any point of M is a measure of the average sectional two dimensional curvature at that point. Hence, the quantity within round parentheses in Eq. (16) could be approximated by $\mathcal{R}/N(N-1)$. This amounts to considering a local average over all the geodesics issuing from a given point and over all the separation vectors orthogonal to them.

However, there is another and *less drastical* approximation to the original Eq. (16): let us consider a *given* geodesic issuing from any point of M, i.e., we keep $\mathrm{d}\mathbf{q}/\mathrm{d}s$ fixed at that point, and let us consider all the possible sectional curvatures obtained by varying only the separation vector ξ; in such a case the sum (19) is replaced by

$$\sum_{a}^{1\ldots N} K^{(2)}_a = \sum_{a}^{1\ldots N} \left(\frac{1}{\|\xi\|^2}\right) R_{ijkl}\xi^i_{(a)} \frac{\mathrm{d}q^j}{\mathrm{d}s} \xi^k_{(a)} \frac{\mathrm{d}q^l}{\mathrm{d}s} = R_{nm} \frac{\mathrm{d}q^n}{\mathrm{d}s} \frac{\mathrm{d}q^m}{\mathrm{d}s} , \tag{20}$$

which is known as the Ricci curvature $K_R(\dot{\mathbf{q}})$ at a point along the direction $\mathrm{d}\mathbf{q}/\mathrm{d}s$, and R_{nm} is the Ricci tensor. This suggests another kind of replacement in Eq. (16), that is $K_R(\dot{\mathbf{q}})/N$. In conclusion, we rewrite Eq. (16) as

$$\frac{1}{2}\frac{\mathrm{d}^2\|\xi\|^2}{\mathrm{d}s^2} + \frac{1}{N}K_R(\dot{\mathbf{q}})\|\xi\|^2 - \left(\frac{\mathrm{d}\|\xi\|}{\mathrm{d}s}\right)^2 = 0\,. \tag{21}$$

As already discussed in [25,26], it is better to use Ricci curvature instead of scalar curvature because the loss of information – going from Eq. (16) to Eq. (21) – is less serious with Ricci curvature.

By the way, since Eq. (21) holds true independently of the choice of the ambient manifold and its metric, it is worth mentioning that if one works with the enlarged configuration space time, equipped with Eisenhart metric, then scalar curvature is *always* identically vanishing, whereas Ricci curvature is not.

We must *always* keep in mind that Eq. (21) is an *approximation* to Eq. (16), on the other hand it has the advantage of describing in closed form the evolution of the norm of the separation vector ξ. By setting $\zeta = \|\xi\|^2$, it can be cast in the form

$$\begin{aligned} &\frac{\mathrm{d}^2\zeta}{\mathrm{d}t^2} - \frac{1}{W}\frac{\mathrm{d}W}{\mathrm{d}t}\frac{\mathrm{d}\zeta}{\mathrm{d}t} + 2\chi\zeta - \frac{1}{2\zeta}\left(\frac{\mathrm{d}\zeta}{\mathrm{d}t}\right)^2 = 0 \\ &\chi = \frac{1}{N}R_{nm}\frac{\mathrm{d}q^n}{\mathrm{d}t}\frac{\mathrm{d}q^m}{\mathrm{d}t}\,. \end{aligned} \tag{22}$$

Details about this equation can be found in [25]. The following results are found: *a*) $\chi < 0$ is a *sufficient* condition to get an exponential growth of $\zeta(t)$, thus to make chaos; *b*) $\chi > 0$ is a *necessary* condition for $\zeta(t)$ to remain bounded, thus for regular behaviour. The latter item deserves particular attention, in fact, $\chi > 0$ is not *sufficient* to ensure the dynamical stability of nearby trajectories, on the contrary a subtle mechanism is generally at work to make chaos also when $\chi > 0$: *parametric resonance*. In other words, the bumpiness of the manifold (M, g_J) can be an effective source of exponential instability of nearby geodesics – thus of chaos – even in presence of positive scalar or Ricci curvature. This can be studied through the following linear second order equation with non constant coefficients[25]

$$\frac{\mathrm{d}^2X}{\mathrm{d}t^2} - \frac{\dot{W}(t)}{W(t)}\frac{\mathrm{d}X}{\mathrm{d}t} + \frac{1}{N}K_R(\dot{\mathbf{q}}(t))X = 0\,, \tag{23}$$

which is derived from Eq. (22). An exponential growth of $X(t)$ implies an exponential growth of $\zeta(t)$ and a bounded evolution of $X(t)$ implies a bounded evolution of $\zeta(t)$.

We have insisted about the reduction of the *exact* Eq. (15) to the *approximate* Eq. (22), or equivalently Eq. (23), because according to a widespread commonplace, borrowed from ergodic theory (but forgetting that here *all* the sectional curvatures must be negative), chaos should be related with negative scalar curvature. But things are not so simple! Let us show why.

Using Eisenhart metric we find

$$\frac{1}{N}K_R(\dot{\mathbf{q}}) = \frac{1}{N}\Delta V(\mathbf{q}) \tag{24}$$

where $V(\mathbf{q})$ is the potential function of a standard Hamiltonian and Δ is the Euclidean Laplacian.

Using Jacobi metric we find

$$\begin{aligned}\frac{1}{N}K_R(\dot{\mathbf{q}}) = &\left[\frac{\Delta V}{2(E-V)} + \frac{4-N}{4(E-V)^2}(\nabla V)^2\right]\delta_{ik}\dot{q}^i\dot{q}^k \\ &+ \frac{N-2}{2(E-V)}\frac{\partial^2 V}{\partial q^i \partial q^k}\dot{q}^i\dot{q}^k + \frac{3(N-2)}{4(E-V)^2}\frac{\partial V}{\partial q^i}\frac{\partial V}{\partial q^k}\dot{q}^i\dot{q}^k\end{aligned} \tag{25}$$

where ∇ is the Euclidean gradient; this can be rewritten as

$$\begin{aligned}\frac{1}{N}K_R(\dot{\mathbf{q}}) = &\frac{\Delta V}{N} + \frac{(\nabla V)^2}{N(E-V)} + \frac{3(N-2)}{4N(E-V)^2}\left(\frac{\mathrm{d}V}{\mathrm{d}t}\right)^2 \\ &+ \frac{N-2}{2N(E-V)}\left(\frac{\mathrm{d}^2 V}{\mathrm{d}t^2}\right).\end{aligned} \tag{26}$$

Now, for convex confining potentials $\Delta V > 0$, as is the case of the FPU model. For the Lennard-Jones potential, and in general for those potentials with an inflection point, it is possible to find $\Delta V < 0$, but this usually happens – say – in the gaseous phase. In condensed matter systems the condition $\Delta V < 0$ is rather exceptional, hence Ricci curvature is always, or almost always, positive in the case of Eisenhart metric, and is positive in the overwhelming majority of points along a geodesic of Jacobi metric. This assertion is based on the results of numerical simulations. For instance, for the already mentioned Hamiltonian models of physical interest, it is an experimental fact that negative Ricci curvature is rarely encountered along a trajectory originated by generic random initial conditions. By inspection of Eq. (26) it is immediately realized that negative contributions can be given by the term containing the Laplacian of V and by the last term. When the Laplacian is alway positive (FPU model, lattice φ^4), or when it is almost always positive, then only the second time derivative of the potential can give some negative contribution, but this is a rare event; moreover, at fixed energy density, the number of times that a negative value of the Ricci curvature is encountered – in a given time lapse – is an inverse function of N[25,26].

In other words, in the framework of Eq. (21) the dominating source of chaos in Hamiltonian flows of physical interest is *parametric resonance due to the curvature fluctuations of the underlying manifold.* Loosely speaking we can affirm that the *bumpiness* of the ambient manifold appears as the relevant geometric property at the origin of chaos rather than some negative curvature property.

In particular, to give a measure of the bumpiness of the manifold, we can compute its average Ricci curvature constraining the computation on the constant energy surface Σ_E as follows

$$\langle k_R \rangle_{\Sigma_E} = \Omega_E^{-1} \int_{\Sigma_E} \mathrm{d}\sigma_E \, k_R \tag{27}$$

where

$$\Omega_E = \int_{\Sigma_E} \mathrm{d}\sigma_E = \int \delta\left(H(\mathbf{q},\mathbf{p}) - E\right) \mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p}$$

and

$$\int_{\Sigma_E} \mathrm{d}\sigma_E \, k_R = \int k_R(\mathbf{q}) \delta\left(H(\mathbf{q},\mathbf{p}) - E\right) \mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p} \,,$$

where $k_R = K_R/N$.

Using Eisenhart metric on $M \times \mathbf{R}^2$, the mean Ricci curvature k_R for the FPU-β model is given by

$$k_R = 2 + \frac{6\mu}{N} \sum_{i=1}^{N} (q_{i+1} - q_i)^2 \,. \tag{28}$$

Notice that here k_R is *always positive*, hence *parametric instability* is the only mechanism to make chaos in FPU, at least in $M \times \mathbf{R}^2$.

In this case, according to Eq. (27) we can compute analytically the average Ricci curvature of $(M \times \mathbf{R}^2, g_E)$ in parametric form[26]

$$\langle k_R \rangle(\epsilon) \begin{cases} \langle k_R \rangle(\theta) = 2 + \dfrac{3}{\theta} \dfrac{D_{-3/2}(\theta)}{D_{-1/2}(\theta)} \\ \epsilon(\theta) = \dfrac{1}{8\mu} \left[\dfrac{3}{\theta^2} + \dfrac{1}{\theta} \dfrac{D_{-3/2}(\theta)}{D_{-1/2}(\theta)} \right] \end{cases} \tag{29}$$

where the D_ν are parabolic cylinder functions. This results is reported here because it provides a good encouragement to pursue the Riemannian description of chaos, in fact notice that in the harmonic limit $\mu = 0$ we find $\langle k_R \rangle(\epsilon) = 2$. The same result is found for the Toda lattice (a nonlinear integrable model), thus suggesting that $\langle k_R \rangle(\epsilon) = \text{const.}$ may be a sufficient condition for integrability. Moreover $\langle k_R \rangle(\epsilon)$ in Eq. (29) is computed *in the thermodynamic limit* ($N \to \infty$) and clearly marks the SST that has been discussed in the preceding Section. Let us look at Fig. 4 where the function (29) is plotted (solid line) for the case $\mu = 0.1$ and compared with time averages $\overline{k_R}$

$$\overline{k_R} = \frac{1}{\tau} \int_0^\tau k_R(\mathbf{q}(t)) \, \mathrm{d}t \,,$$

computed along numerical trajectories obtained at different ϵ and at $N = 128$, 512. Not only the agreement is strikingly good, but it shows two asymptotic behaviours: at low ϵ we observe a weak deviation of $\langle k_R \rangle(\epsilon)$ from constancy – which can be naturally interpreted as a mark of weak chaos –, whereas at

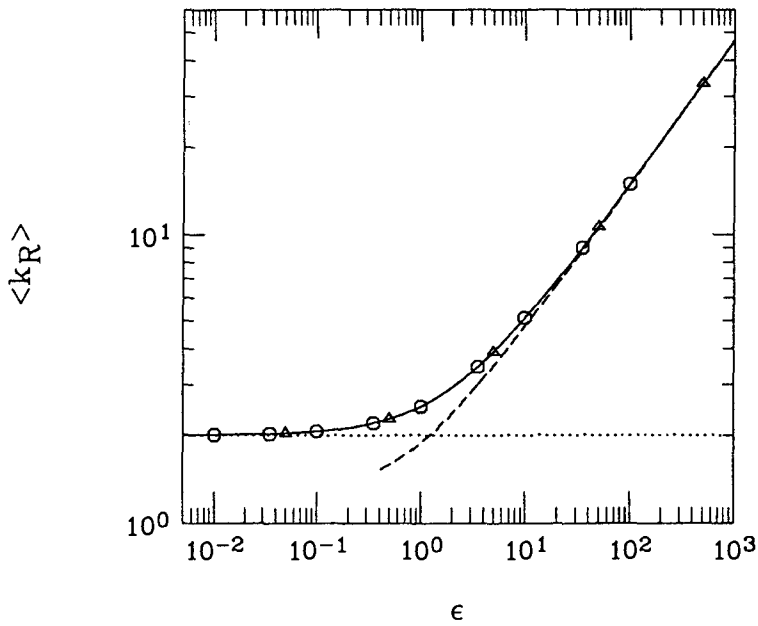


Fig. 4. FPU model. The average Ricci curvature (solid line) of the enlarged configuration space time equipped with Eisenhart metric is computed on the constant energy surface Σ_E and plotted as a function of the energy density ϵ [Eq. (30)]. The same quantity is numerically computed along trajectories obtained with random initial conditions at $N = 128$ (circles) and at $N = 512$ (triangles). The horizontal line, $\langle k_R \rangle = 2$, refers to a collection of harmonic oscillators ($\mu =$ limit). The SST is clearly marked

high ϵ we observe a faster increase of $\langle k_R \rangle(\epsilon)$ which is naturally interpreted as a mark of strong chaos. The crossover is here in excellent agreement with that observed in the scaling of $\lambda_1(\epsilon)$.

Notice that, at any given ϵ, it is the knowledge of $\mathrm{d}\langle k_R \rangle/\mathrm{d}\epsilon$ that indicates whether chaos is weak or strong; this is a first simple step in finding geometric quantities free of problems when N is changed. It is also possible to make further improvements of the approximations that bring Eq. (16) to Eq. (21). It would take too much place here even to give a flavour of how such improvement can be done, anyway it is now possible to give a reliable measure of the strenght of chaos by purely geometric quantities and so to reproduce analytically[30] the numerical result $\lambda_1(\epsilon)$ obtained for the FPU model and reported in Fig. 2.

Finally, we want to mention that the arguments discussed in this Section fit in a consistent scenario based on the comparison among analytic and numeric results; this is just the infancy of this approach, at least with the aims of theoretical physics rather than of mathematics.

Having elucidated some simple but basic points with the more friendly models of condensed matter physics, we can conclude this contribution with a glance at the gravitational N-body models.

There are pioneering papers[31,32] concerning a Riemannian approach to the description of the mixing properties of the gravitational N-body systems, which have the great merit of having resumed the line of thought opened by Krylov half a century ago and almost forgotten by physics. However, the problems raised above about the delicate passage from Eq. (16) to Eq. (21) still apply to the self-gravitating N-bodies; with the aid of computer simulations one finds here again[29] that for random and almost virialized initial conditions the Ricci curvature of M is always positive and that parametric instability is the actual mechanism that makes chaotic the gravitational systems too. Thus we cannot properly use neither the language nor the results of rigorous ergodic theory. Moreover a reliable estimate – based on geometric quantities – of the dynamical instability time scale requires those refinements that have been developed in [30]. It is perhaps worth keeping in mind also that if the phase space of the gravitational systems is not like that of an Anosov flow, but has – loosely speaking – a "more complex" structure, then it might happen, as we have seen for the FPU model, that the dynamical instability time scale and whatever relaxation time scale one may define are probably related in a very complicate and subtle manner.

References

[1] B. Mandelbrot, Ann. Math. Stat. 33, 1021 (1962)
[2] R.H. Miller, Ap. J. 140, 250 (1964); J. Comp. Phys.8, 449 (1971)
[3] E. Fermi, J. Pasta, and S. Ulam, in *Collected Papers of Enrico Fermi*, edited by E. Segré (University of Chicago, Chicago, 1965), Vol. 2, p. 978
[4] A. N. Kolmogorov, Dokl. Akad. Nauk SSSR 98, 527 (1954)
[5] V. I. Arnold, Russ. Math. Surv. 18, 9 (1963)
[6] J. Moser, Nachr. Akad. Wiss. Goettingen Math. Phys. Kl.2 1, 1 (1962)
[7] H. Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste* (Blanchard, Paris, 1987), Vol. 3, p. 389; E. Fermi, Nuovo Cimento 25, 267 (1923); 26, 105 (1923)
[8] L. Chierchia, and G. Gallavotti, Nuovo Cimento B67, 277 (1982); G. Benettin, L. Galgani, A. Giorgilli, and J. M. Strelcyn, Nuovo Cimento B79, 201 (1984)
[9] N. N. Nekhoroshev, Funct. Anal. Appl. 5, 338 (1971); Russ. Math. Surv. 32, 1 (1977)
[10] G. Benettin, and G. Gallavotti, J. Stat. Phys. 44, 293 (1986)
G. Benettin, L. Galgani, and A. Giorgilli, Celest. Mech. 37, 1 (1985)
P. Lochak, Phys. Lett. A143, 39 (1990); "Stabilité en temps exponentielle des systèmes hamiltoniens proches de systèmes intégrables: résonances et orbites fermées", D.M.I., Ecole Normale Supérieure, Paris, (1990)
[11] L. Casetti, *An efficient symplectic algorithm for numerical simulations of Hamiltonian flows*, Phys. Scr., (1993) submitted
[12] N. J. Zabusky, M. D. Kruskal, Phys. Rev. Lett. 15, 240 (1965)

[13] F. M. Izrailev, B. V. Chirikov, Dokl. Akad. Nauk SSSR 166, 57 (1966) [Sov. Phys. - Dokl. 11, 30 (1966)]
[14] B. Callegari, M. Carotta, C. Ferrario, G. Lo Vecchio, L. Galgani, Nuovo Cimento B54, 463 (1979)
G. Benettin, G. Lo Vecchio, A. Tenenbaum, Phys. Rev. A22, 1709 (1980)
G. Benettin, A. Tenenbaum, Phys. Rev. A28, 3020 (1983)
[15] We quote for all the first pioneering paper on this subject: P. Bocchieri, A. Scotti, B. Bearzi, A. Loinger, Phys. Rev. A2, 2013 (1970)
[16] M. Pettini, M. Landolfi, Phys. Rev. A41, 768 (1990)
[17] M. Pettini, M. Cerruti-Sola, Phys. Rev. A44, 975 (1991)
[18] R. Livi, M. Pettini, M. Sparpaglione, S. Ruffo, A. Vulpiani, Phys. Rev. A31, 1039 (1985); R. Livi, M. Pettini, S. Ruffo, A. Vulpiani, Phys. Rev. A31, 2740 (1985)
[19] R. Livi, M. Pettini, S. Ruffo, A. Vulpiani, J. Stat. Phys. 48, 539 (1987)
[20] L. Casetti, Thesis, Dept. of Physics, University of Florence, 1993
[21] P. Butera, G. Caravati, Phys. Rev. A36, 962 (1987)
[22] P. Foggi, V. Schettino, Riv. Nuovo Cim. 15, n°7, (1992)
[23] N.S. Krylov, Thesis, 1942, reprinted in: *Works on the Foundations of Statistical Physics*, Princeton University Press, (Princeton, N.J. 1979)
[24] See for a review: Ya.G. Sinai, *Dynamical Systems II*, Springer-Verlag, 1989
[25] M. Pettini, Phys. Rev. E47, 828 (1993)
[26] L. Casetti, M. Pettini, Phys. Rev. E, (1993), in press
[27] J. Guckenheimer, P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcation of Vector Fields*, Springer-Verlag, 1983
[28] E.T. Whittaker, *A Treatise on Analytical Dynamics of Particles and Rigid Bodies*, Cambridge University Press, p.419, 1937
[29] M. Cerruti-Sola, M. Pettini, Astron. & Astrophys., (1993) submitted
[30] L. Casetti, R. Livi, M. Pettini, Phys. Rev. Lett., (1993) submitted
[31] V.G. Gurzadyan, G.K. Savvidy, Astron.& Astrophys. 160, 203 (1986)
V.G. Gurzadyan, A.A. Kocharyan, Astrophys. & Space Sci. 133, 253 (1987)
V.G. Gurzadyan, A.A. Kocharyan, Astrophys. & Space Sci. 135, 307 (1987)
[32] H.E. Kandrup, Astrophys. J. 364, 420 (1990)

Numerical Exploration of the Circular Billiard with Gravity

Avram HAYLI [1] & Augustin VIDOVIĆ [2]

[1] Observatoire de Lyon, 69230 Saint-Genis-Laval, FRANCE
[2] Université Claude-Bernard, Lyon 1, FRANCE

Abstract. The properties of the billiard vary with the energy E of the bouncing ball. These are explored through the surface of section. The billiard is almost integrable for very small or very large values of E. It looks ergodic in a range of values of E. The main families of periodic orbits are shown.

1. Introduction

Let Ω be an open region of the horizontal plane $\mathbb{R}^2$ whose boundary is a strictly convex oriented closed curve $\partial\Omega$ of class C^k, $k \geq 2$. A billiard in Ω is the dynamical system defined by the free motion of a particle in the interior of this enclosure with specular bounces on the boundary. It is a Hamiltonian system with two degrees of freedom.

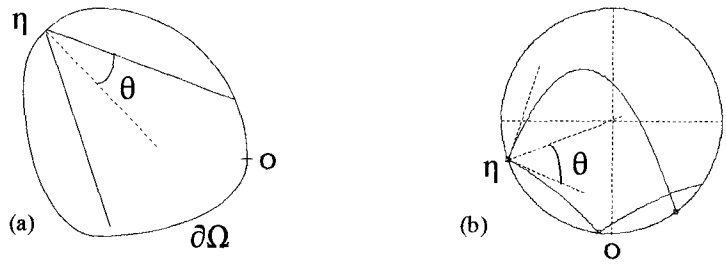


Fig. 1. (a) The horizontal billiard. (b) The circular billiard with gravity

A convenient way to explore the properties of a billiard is through the surface of section. We look at the successive bounces on the boundary. Let η be the curvilinear abscissa, measured along $\partial\Omega$ from the point of impact to an arbitrarily chosen origin O, and θ the oriented angle measured from the normal to $\partial\Omega$ at the point of impact to the incident trajectory as in Fig. 1a. We assume that $\partial\Omega$ has unit length. We take as coordinates in the surface

of section η and $S = \sin\theta$. It is clear that $0 \leq \eta \leq 1$ and $|S| \leq 1$. The surface of section is then identified with the rectangle

$$\Pi = \{(\eta, S) \in \mathbb{R}^2 \quad 0 \leq \eta \leq 1\ ,\ -1 < S < 1\}. \tag{1}$$

The motion in the billiard is described by a map $T : \Pi \to \Pi$ of the surface of section onto itself that preserves the Lebesgue measure $\mathrm{d}\eta\,\mathrm{d}s$.

A billiard is said to be integrable if the surface of section is filled with invariant curves. Only two billiards are known to be integrable: the one inside a circle, for which the invariant curves are the lines $S = \text{const.}$, or inside an ellipse. A billiard is said to be ergodic if no invariant curves exist. The usual situation is that of a coexistence of regions with invariant curves and regions without such curves. Douady (1982) has shown that for $k > 6$ there is a set of invariant curves of positive measure close to $S = 1$ or to $S = -1$. He also conjectured that this is true for $k > 4$. This means that such billiards are not ergodic. If $k = 1$ one cannot rule out the existence of ergodic billiards as has been shown by Hayli & Dumont (1986) on the basis of numerical evidence. The same seems true if $k = 2$. It is clear however that no certainty could be attained by this approach, as can be seen by comparing Robnik (1983) & Hayli et al. (1987).

2. The Circular Billiard on an Inclined Table

The billiard we study now is circular and lies on a table which is plane but not horizontal. In other words we shall explore the circular billiard in the presence of gravity. Other billiards in a gravitational field have already been explored. See for example the billiard in a wedge studied by Lehtihet & Miller (1986). The trajectory of the particle between two elastic bounces is an arc of parabola (Fig. 1b). By contrast with the horizontal billiard the velocity of the particle plays now a role. The dynamical system depends on several parameters. These are the radius of the circle, the angle of inclination of the table, the mass and the energy of the particle. It is easily seen that one parameter is enough to describe the system. We take the radius of the circle, the acceleration of gravity and the mass of the particle equal to one. Accordingly we shall vary only the energy of the particle. We give in the following some preliminary results of the numerical exploration. Theoretical results will be given elsewhere.

2.1 The Surface of Section

Figure 2 summarizes the main features in the surface of section for E ranging from 0 to ∞. The origin is the lowest point in the billiard and η is reckoned from -0.5 to 0.5. It is clear that when $E < 2$ the surface of section is a rectangle smaller than Π. In this case only the relevant rectangle is represented.

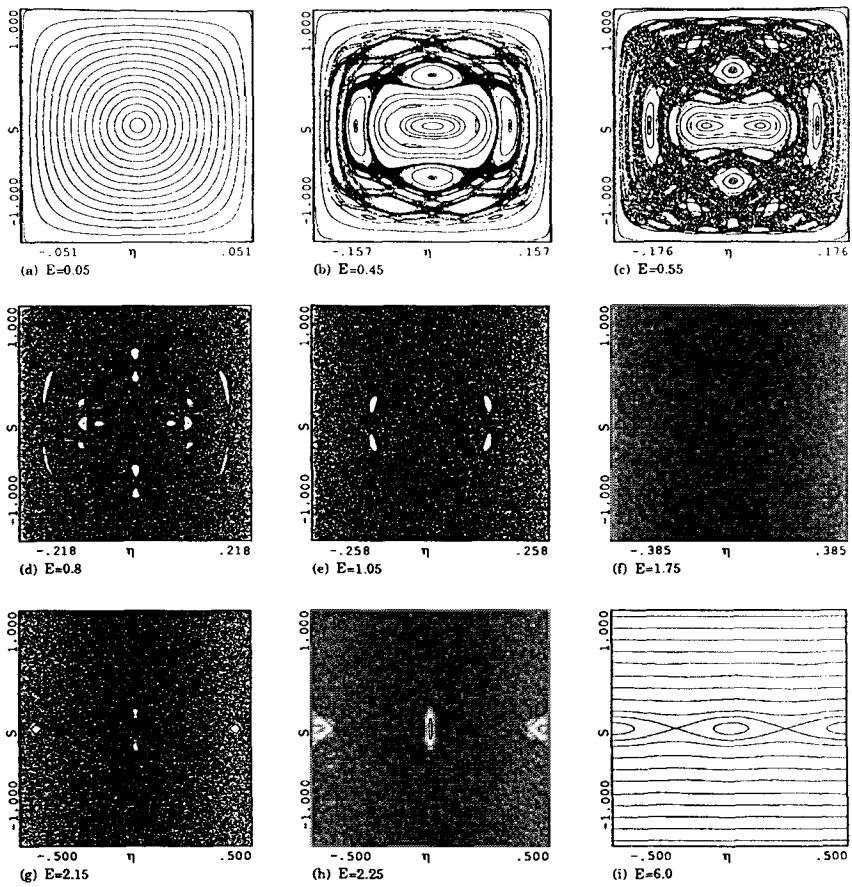


Fig. 2. The surface of section for different values of the energy E

For small values of E the surface of section seems to be filled with invariant curves. Fig. 2a is for $E = 0.05$; the billiard looks integrable. In fact one could probably find a very thin ergodic strip outside the curves. The general aspect changes drastically for $E > 0.25$. Fig. 2b-c are for $E = 0.45$ and $E = 0.55$. The surface of section has the familiar aspect met for horizontal billiards. Islands develop around invariant points corresponding to stable periodic orbits. The ergodic region has greater extension with increasing E. The billiard looks ergodic from about $E = 1.15$ to E slightly greater than 2. However careful examination shows that this is not quite true. Very small islands are observed for $E = 1.2$, beyond the classical doubling cascades, as has been noticed by MacKay (1982) in other circumstances. We think that the billiard is probably ergodic for E ranging from 1.75 to about 2. Let us point out that for $E = 1.75$ there is a trajectory starting from $\eta = 0$ which is osculating to the boundary. Islands reappear for E slightly greater than 2. The size of the region with

invariant curves increases when E becomes larger. Although the billiard is integrable only for $E = \infty$ we observe on Fig. 2i that it looks integrable already for $E = 6$.

2.2 The Main Families of Stable Periodic Orbits

These are shown in Fig. 3. Orbits of family I are in Fig. 3a-d. For energies smaller than 0.5 there is a one-periodic rectilinear stable orbit starting from $\eta = 0$ normal to the billiard. It corresponds to the invariant point at the centre of Fig. 2a. For $E = 0.5$ we observe a bifurcation. A stable symmetric two-periodic parabolic orbit as in Fig. 3b appears. An orbit of this type corresponds to the two invariant points in the central part of Fig. 2c. For $E = 1$ this orbit splits into two identical orbits described in opposite directions. This is shown in Fig. 3c. These orbits correspond to the invariant points inside the islands in Fig. 2e. Orbits of this type remain stable until $E \simeq 1.17678$. It is remarkable that the point of impact A does not vary for $1 < E < 1.17678$; namely one has $\eta(A) = 1/8$. For $E \simeq 1.17678$ these orbits become unstable and gives by bifurcation a 4-periodic orbit of the type shown in Fig. 3d. This orbit in turn becomes unstable for $E \simeq 1.19882$ and the cascade of bifurcations continues as in Hayli et al. (1987).

Orbits of family II are shown in Fig. 3e-h. They do not exist for $E < 0.25$. Orbits of the type in Fig. 3e are 4-periodic and symmetric; they start normal to the billiard. Such an orbit corresponds to the invariant points inside the four islands of Fig. 2b. For $E = 0.75$ it is replaced by two identical symmetric orbits described in opposite directions like the one shown in Fig. 3f. Orbits of this type remain stable until $E \simeq 0.84003$. Again it is remarkable that the point of impact B remains the same, namely one has $\eta(B) = 1/6$. For $E \simeq 0.84003$ this orbit becomes unstable and gives by bifurcation an 8-periodic symmetric orbit of the type shown in Fig. 3g. For $E \simeq 0.84390$ the 8-periodic orbit becomes asymmetric as in Fig. 3h.

Then for $E \simeq 0.84963$ this orbit becomes unstable and gives through bifurcation a 16-periodic orbit (not shown).

Orbits of family III are shown in Fig. 3i-k. They appear for $E \simeq 2.0099$ as soon as the billiard has ceased to look ergodic. The first type is 4-periodic and asymmetric. It becomes symmetric for $E \simeq 2.03078$. For $E = 2.25$ we observe an inverse bifurcation and the 2-periodic orbit along the diameter becomes stable. It remains so for all greater values of E.

One orbit of family IV is shown in Fig. 3l. These 3-periodic orbits have a parabolic part and a rectilinear part. The type in Fig. 3l is associated with the invariant points inside the three small islands like those in Fig. 2b. A bifurcation of this type of orbit has been observed with doubling period.

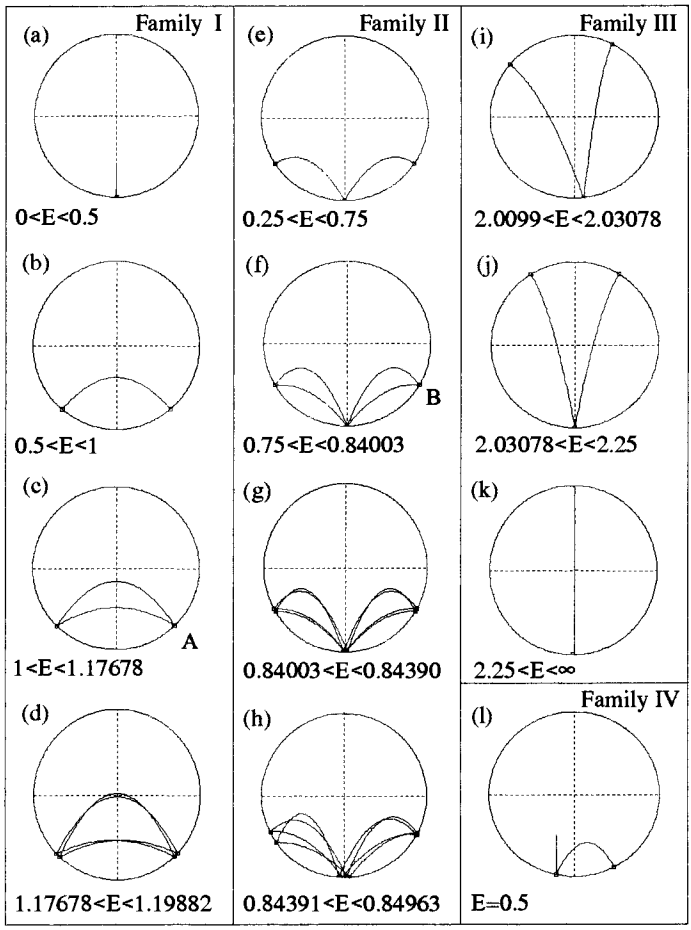


Fig. 3. The main families of stable periodic orbits

3. Conclusion

The billiard inside an inclined circle might be considered in some way as a perturbation of the horizontal one. It is almost integrable for small or large energies of the bouncing ball. It looks ergodic for some intermediate energies. We think that ergodicity has to do with the occurrence of osculating trajectories to the boundary. Several families of relatively simple periodic orbits have been found. Only few of them were presented here.

More work on the subject is in progress.

References

Douady R., 1982, Thèse de 3ème cycle Paris VII
Hayli A., Dumont Th., 1986, Celes. Mech. 38, 23
Hayli A., Dumont Th., Strelcyn J.M., Moulin-Ollagnier J., 1987, J. Phys. A: Math. Gen. 20, 3237
Lehtihet H.E., Miller B.N., 1986, Physica 21D, 93
MacKay R.S., 1987, Phys. Lett. 87A, 321
Robnik M., 1983, J. Phys. A: Math. Gen. 16, 3971

Ergodicity and Mixing in Gravitating Systems

V.A. ANTONOV [1], L.P. OSSIPKOV [2]

[1] Institute for Theoretical Astronomy, St Petersburg, Russia
[2] Institute of Computational Mathematics and of Processes Control, St Petersburg State University, Russia

Abstract. When considering the ergodic problem for gravitating systems of N point masses we have to take into account the circumstance that for large enough N (i.e. for galaxies of the most usual types) the systems are practically collisionless for time scales not exceeding the Hubble time. So it is necessary to follow phase trajectories not only in the $6N$-D phase space but in the 6-D phase space. Studying the $6N$-D space will give no new information for completely collisionless systems. We suggest a classification of various kinds of mixing in collisionless gravitating systems based mainly on degree of non-stationarity of the smoothed gravitational field (Antonov, Nuritdinov, & Ossipkov 1973). We distinguish a compulsive mixing in violently non-stationary systems, a quasi-diffusion mixing in weakly non-stationary systems, a divergent mixing which is connected with an exponential divergence of initially close trajectories for steady-state non-integrable systems, and a circulation mixing that will be in integrable systems in the case of dependence of circulation frequencies on values of isolating integrals of motion. The general property of collisionless mixings of any kind is the increase of the quasi-entropy, i.e., an integral of any convex function of the coarse-grained distribution function (Antonov 1963; Antonov et al. 1973; Tremaine et al. 1980). So we can show that some evolutionary ways are impossible for isolated collisionless systems. For example spherical systems cannot evolve along the sequence of polytropic indices (Antonov 1990). Any mixing cannot increase the maximal value of the distribution function.

References

Antonov V.A., 1963, Thesis, Leningrad Univ.
Antonov V.A., 1990, Probl. Cel. Mech. and Stellar Dynamics, ed. T.B.Omarov, Nauka, Alma-Ata, p.131
Antonov V.A., Nuritdinov S.N., Ossipkov L.P., 1973, Dynamics of Galaxies and Star Clusters, ed.T.B.Omarov, Nauka, Alma-Ata, p.55
Tremaine S., Hénon M., Lynden-Bell D., 1980, MNRAS 218, 285

3. Computer Simulations and Mappings

Chaotic Itineracy and Clustered Motion in Globally Coupled Symplectic Map System

Tetsuro KONISHI [1] & Kunihiko KANEKO [2]

[1] Dept. of Physics, School of Science, Nagoya Univ., Nagoya 464-01, Japan
[2] Dept. of Pure and Applied Sciences, College of Arts and Sciences, Univ. of Tokyo, Komaba, Meguro-ku, Tokyo 153, Japan

Abstract. N-particle systems of 1-D globally coupled maps are numerically investigated to show the formation of spatial order in chaotic dynamics of Hamiltonian systems. The system itinerates between ordered and disordered states by its deterministic dynamics. Phase space structure and existence of long-time tail reveals that the ordered state is supported by hierarchical motion around ruins of regular orbits (KAM tori and islands).

1. Introduction

Stellar systems show interesting behaviours in that, while they are governed by conservative dynamics without dissipation, they avoid thermalisation and form spatial structures, such as elliptical galaxies and globular clusters. The fundamental interaction between each star is completely known, but we have not yet understood the origin of the entire structure. For example, it is difficult to understand the triaxial shapes of elliptical galaxies by superposition of regular orbits, since they are born and grown through complicated interactions among stars. Thus we need a new theory to explain spatial order.

Chaotic dynamical systems shed a new light to structure formation. Various spatial structures appear in various physical systems. Some of them are familiar to us, such as crystals, linear waves, solitons. They are regular, periodic or stationary structures. There are many other dynamic patterns such as in turbulence, vortices, liquids, microclusters. To describe non-stationary and dynamical order, we need power of chaos.

Long-time behaviour of chaotic dynamics in conservative systems are usually considered to be of two classes. One is thermal equilibrium, which is spatially uniform (translational symmetry), and shows exponential relaxation to the equilibrium. The other is $1/f$-type fluctuations, or long-time tail, mostly found in small systems like the standard map.

There is another kind of conservative chaotic systems, which forms non-uniform spatial structure. Glassy systems[5], hydrogen bond network in liquid

water[6], microclusters[7], and self-gravitating systems. In this paper we describe, with the use of a globally coupled map system, a simple model of such a class of dynamical systems which develops spatial structure with Hamiltonian dynamics.

2. Model

Our model is a coupled map lattice with symplectic condition[1]. Coupled map lattices are defined on discrete time and used in various fields for its numerical efficiency. In particular they are suitable for numerical investigation of phenomena in which long-time behaviour is important.

In our model we have N particles on the 1-dimensional unit circle (hence our space is compact). The state of each particle i is defined by a pair of real number (x_i, p_i). x_i is its phase (position) with $0 \leq x_i < 1$ and p_i is its conjugate momentum. The temporal evolution rule is defined as $(x_i, p_i) \mapsto (x_i', p_i')$, $i = 1, 2, \ldots, N$:

$$\begin{aligned} p_i' &= p_i + \frac{K}{2\pi\sqrt{N-1}} \sum_{j=1}^{N} \sin 2\pi(x_j - x_i) \,, \qquad K > 0 \,. \\ x_i' &= x_i + p_i' \,, \end{aligned} \tag{1}$$

Since $K > 0$, the interaction term $\left(K/2\pi\sqrt{N-1}\right) \sin 2\pi(x_j - x_i)$ between two particles i and j is *attractive*. The interaction is long range and tends to zero at close 2-body encounters.

The model Eq. (1) satisfies the symplectic condition: $\sum_{i=1}^{N} \mathrm{d}x_i \wedge \mathrm{d}p_i = \sum_{i=1}^{N} \mathrm{d}x_i' \wedge \mathrm{d}p_i'$, so that the model can be regarded as a Poincaré mapping for a Hamiltonian system with $N+1$ degrees of freedom. Another interpretation of our model is a use of a "kicked" Hamiltonian, as in the standard mapping. Note that the total energy is not a constant of motion. The entire phase space of the model corresponds to an energy surface of an energy conserving system. The total momentum $\sum_{j=1}^{N} p_j$ is a constant of motion.

3. Clustered State of Particles

Figures 1 and 2 are typical temporal evolution of the model. In Fig. 1 particles, initially spread on the unit circle, gradually assemble and form a cluster. These figures are taken for the same parameter $K = 0.1$ and their only differences are the initial conditions. For clustered motion the initial momenta are small (say $|p_i| < 0.2$) and for non-clustered motion the initial momenta are large.

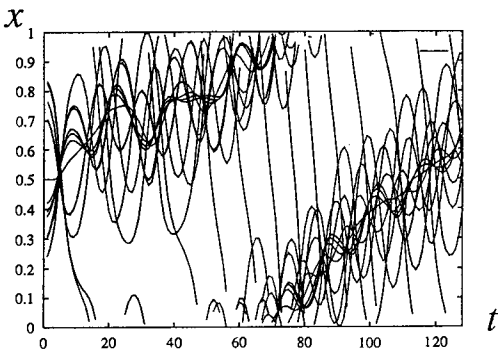


Fig. 1. Clustered motion. $N = 12$, $K = 0.1$. Initial momenta are all 0

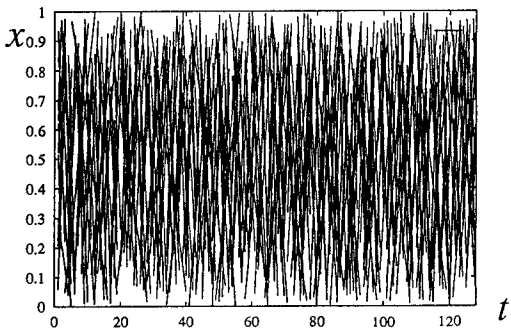


Fig. 2. Non-clustered motion. $N = 12$, $K = 0.1$. Initial momenta are random over $[0, 1]$

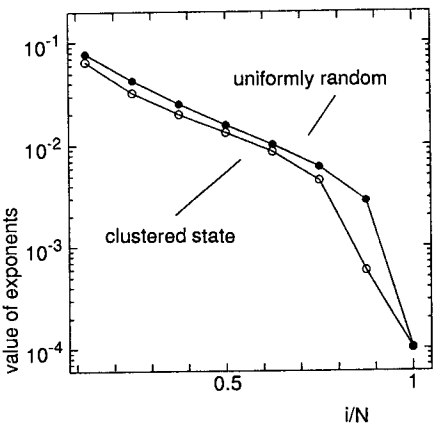


Fig. 3. Lyapunov spectra of clustered and uniform random state

4. Lyapunov spectra

Figure 3 shows the Lyapunov spectra of the model Eq. (1) for ordered (clustered) and disordered (non-clustered) states. Each spectrum is calculated by averaging over 100 orbits for each states. From the figure we see that clus-

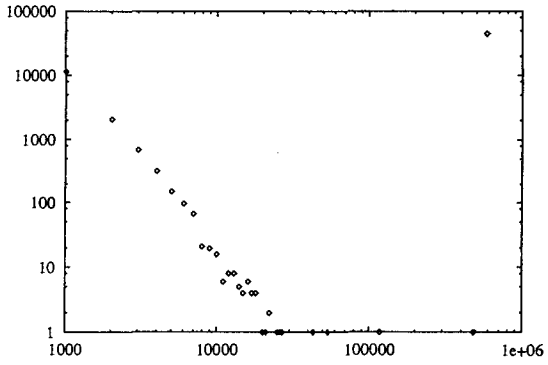


Fig. 4. Lifetime distribution of clustered state: log-log

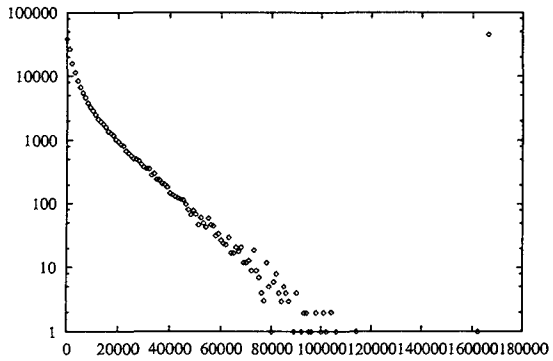


Fig. 5. Lifetime distribution of non-clustered state: semi-log

tered motion is chaotic, and we have that two different chaotic seas coexist in phase space. The existence of different chaotic seas implies that these seas are topologically connected, since KAM tori (if any), which is only N dimensional manifold, cannot divide the $2N$-dimensional phase space into disconnected parts if $N > 1$. And we expect that these chaotic seas are dynamically connected. Indeed, clusters once formed dissolve into non-clustered state, and vice versa.

This kind of successive transition among various states by internal dynamics is called "chaotic itinerancy" and is attracting attention in nonlinear optics, neural systems, etc.[4]

Figures 4 and 5 represent the lifetime distribution of clustered and non-clustered states taken from a single time series with $N = 4$, $K = 0.2$. The distributions are fitted as

$$P(t) \propto \begin{cases} t^{-c} & \text{for clustered state,} \\ \exp(-\alpha t) & \text{for non-clustered state.} \end{cases} \tag{2}$$

An exponential distribution in non-clustered state implies that the temporal correlation is negligible and the system must find entrances to the clustered

state "by chance". A power law distribution in clustered state means that there is no typical time scale in the ordered state. It is natural to interpret this power law as a result of a hierarchical phase space structure[1].

5. Summary and Discussion

We have shown that Hamiltonian systems can form spatial structures with the use of chaotic dynamics. The structure is deeply connected with a hierarchical structure of phase space. Combined with other studies[1], we give a summary in the following table:

Table 1. Properties of two chaotic states. Summary.

state	clustered	non-clustered
real space structure	cluster of particles	none
phase space structure	ruins of KAM tori, islands	none
temporal correlation	long (power law)	short (exponential)
resident time distribution	power	exponential

Although our model is not directly connected to astrophysical objects, we think it is conceptually important, since it gives a concrete example of formation of non-symmetric order with chaotic dynamics, which is the case of elliptical galaxies[2]. More studies are needed from various aspects. Thermodynamical approach to this model is successfully described in [3]. Chaos is not just a mess, it is a rich source of various active orders.

Acknowledgements. We would like to thank S. Inagaki and N. Gouda for introducing us to the wonderful world of astrophysics. We are grateful to the National Institute for Fusion Science at Nagoya for computational facility of FACOM VP200E, VP200 and M380. The present research is partially supported by Grant-in-Aids for Scientific Research from the Ministry of Education, Science, and Culture of Japan.

References

[1] Konishi T., Kaneko K., 1992, J. Phys. A 25, 6283
[2] Gouda N., this proceedings
[3] Inagaki S., this proceedings
[4] Ikeda K., Matsumoto K., Ohtsuka K., 1989, Prog. Theor. Phys. Suppl. 99, 295; Tsuda I., 1990, in Neurocomputers and Attention, A.V. Holden & V. I. Kryukov (eds.), Manchester Univ. Press
[5] Shinjo K., 1989, Phys. Rev. B40, 9167
[6] Ohmine I., Tanaka H., 1990, J. Chem. Phys. 93, 8138
[7] Sawada S., 1987, in Microclusters, S. Sugano et.al. (eds.), Springer, Berlin, p. 211

Lyapunov Analysis of Stable Chaos in Self-Gravitating Many Body Systems

Naoteru GOUDA [1], Toshio TSUCHIYA [2], & Tetsuro KONISHI [3]

[1] Department of Earth and Space Science, Osaka University, Toyonaka 560, Japan
[2] Department of Physics, Kyoto University, Kyoto 606, Japan
[3] Department of Physics, Nagoya University, Nagoya 464-01, Japan

Abstract. Elliptical galaxies seem to have two *inconsistent* characters in their dynamical structures: triaxial velocity dispersions of the stars suggest that the elliptical galaxies consist of regular stellar orbits, whereas nonlinear interaction in self-gravitating many-body systems makes the elliptical galaxies chaotic. A new type of chaos, "stable chaos", in which orbits do not change significantly although they have orbital instability (positive Lyapunov exponent), is expected to give the key to clarify the dynamical structure of elliptical galaxies. As a first step, we examine whether the stable chaos appears or not in one-dimensional self-gravitating mass sheet systems.

1. Introduction

Elliptical galaxies are considered to be collisionless self-gravitating many-body systems because their two-body relaxation time is much longer than the age of the universe. And they are also considered to stay today at the quasi-equilibrium state because of the universality of their density profiles and of their morphologies. However we do not understand the relaxation mechanism and the dynamical structures in elliptical galaxies.

One may consider the dynamical structure of elliptical galaxies in equilibrium as follows: the orbits of most of the stars are regular, that is, the one-body distribution function has three isolating integrals. This is because some observations suggest that many elliptical galaxies are triaxial ellipsoids and their shape might be supported by the anisotropy of the velocity dispersion (Davies et al. 1983). If the system is chaotic and so the one-body distribution function depends on only the total energy E, then the velocity dispersion in the system is isotropic. Thus the observed (triaxial) anisotropic velocity dispersions suggest the existence of three isolating integrals and hence of *regular* systems.

On the other hand, it is natural to think that almost all orbits in many-body self-gravitating systems are chaotic, since the interaction between each star is nonlinear. Moreover elliptical galaxies are considered to approach an

equilibrium state from their initial state by a violent change of the mean potential which is associated with the chaotic stellar motions. Thus it is difficult to expect that most of the chaotic orbits in galaxies turn to the regular ones through complicated interactions of nonlinear self-gravity, though the present equilibrium states are thought to consist of regular orbits as mentioned above. Then how can we combine these two opposite characters of dynamical structures in elliptical galaxies; those are, regular orbits which support the anisotropic velocity dispersions, and chaotic orbits which might be expected in many-body self-gravitating systems and result in the violent relaxation? What is the dynamical structures of today's elliptical galaxies?

Recently, a new type of chaos besides the familiar ergodic chaos has been found in some Hamiltonian systems (Konishi & Kaneko 1992). They showed that there can exist several different chaotic seas besides the equilibrium state. The chaotic seas are topologically connected but it takes time to go around the whole phase space, and such systems appear to lack ergodicity if one observes them within a limited time scale. Then one can observe that the system is chaotic and at the same time it holds a non-isotropic shape. This type of chaos is called "*stable chaos*".

One of the mechanisms causing stable chaos is that the orbit stays very near the KAM torus for a long time. We expect that some self-gravitating many-body systems may consist of such stable chaotic orbits. Then, though each orbit itself is chaotic, the system might appear approximately to have isolating integrals because each orbit goes around near a torus. We expect that the stable chaos may solve the problem of the inconsistency of the two opposite characters of dynamical structures in elliptical galaxies.

The purpose in this paper is that we examine whether the stable chaos appears really in many-body self-gravitating systems or not.

2. Stable Chaos

We will briefly comment on the stable chaos, which is a key concept to understand dynamical characters of elliptical galaxies.

We generally understand temporal evolution of systems in terms of dynamical system theory and ergodic theory, where properties are usually defined for infinite time (e.g., ergodic property, Lyapunov numbers, and so on). On the other hand, when we are interested in properties of the systems within finite time (e.g., shorter than the age of the universe), we need new concepts in addition to ordinary dynamical system theory.

One example of such a concept is "stagnant motion", which is found for motion around KAM tori in generic Hamiltonian systems. The stagnant layer is defined to be the transit region between the KAM torus and the chaotic sea (Aizawa et al. 1989). In this layer, it is found that the orbits go around the KAM torus for a very long time. We call this motion the stagnant motion.

When stagnant motion exists, the system shows chaotic motion (positive Lyapunov exponent) and at the same time the orbit itself does not change significantly since it is stuck to a KAM torus. Milani & Nobili (1992) found this kind of chaos in the motions of planets and call this chaos "stable chaos". On the other hand, the familiar chaos which is already well known is called "ergodic chaos" because the physical quantities measured as time averages are roughly equal to averages over the phase space.

The orbits in the stable chaos will move to the ergodic chaos region and they will return back to the stagnant layer. The size of the KAM torus is scale-invariant, that is, the distribution of the KAM torus is fractal. This fact results in that the distribution of the pausing time T, which is the staying time at the state of the stable chaos, obeys the power law $P(T) \propto T^{-\alpha}$. Here $P(T)$ is the distribution of the pausing time T and α is a constant. This means that there is no characteristic time scale for the transition from the stable chaos to the ergodic chaos. On the other hand, the pausing time distribution for the ergodic chaos obeys the exponential form $P(T) \propto \exp(-\beta T)$. Here β is a constant. From the above facts, we can find that the time-dependent Lyapunov number, which is defined by

$$\lambda(t) = \frac{1}{t} \log ||\delta \mathbf{x}(t)|| \,, \tag{1}$$

changes as time goes if the transition between stable and ergodic chaos occurs in the system. Here $\delta \mathbf{x}(t)$ is the linear deviation of an orbit at the time t.

3. Mass Sheet Models

Next we will examine whether the stable chaos really appears in self-gravitating systems. As a first step, we analyse the one-dimensional system, in which infinite mass sheets move along only the x-axis (Sakagami & Gouda 1991). The Hamiltonian in this system is given as follows:

$$H = \frac{m}{2} \sum_{i=1}^{N} v_i^2 + 2\pi G m^2 \sum_{i>j} |x_i - x_j| \,, \tag{2}$$

where m, x_i and v_i are the mass per unit area, the location and velocity of the i^{th} sheet, respectively. Of course, this kind of one-dimensional model is far from reality. However we can pursue the evolution of the systems with remarkably high precision. Indeed, the motion of each sheet between encounters with other sheets has a uniform acceleration. Thus we can pursue the evolution of the systems with high accuracy by joining the analytic solutions at the moments when an encounter takes place.

As for the initial conditions, we consider two cases. One is a random distribution in phase space and the other is that all sheets have the same energy. We considered the cases for the total number of mass sheets $N = 4$

to 50. For these initial conditions, it is well known that a core-halo structure appears and the distribution function at the quasi-equilibrium state does not correspond to the Lynden-Bell distribution (Yamashiro et al. 1992).

4. Numerical Results

We will briefly summarise our numerical results for the time-dependent Lyapunov numbers in the mass sheet models. The results are almost the same for the both initial conditions. As for the larger time-dependent Lyapunov numbers, they approach the constant value rather monotonously after an enough long time. The smallest time-dependent Lyapunov number and the next to the smallest one correspond to the perturbations with respect to the total energy and the total momentum conservation, respectively. For these perturbations, the time-dependent Lyapunov numbers are zero and then we found numerically the time-dependent Lyapunov numbers evolve proportionally to the inverse of the time as theoretically expected (Goldhirsch et al. 1987). So we can conclude that we can get the time-dependent Lyapunov spectrum numerically with high accuracy after enough long time.

On the other hand, as for the smaller time-dependent Lyapunov numbers, we found that they do not decrease proportionally to the inverse of the time, but they change by an order of magnitude and do not stably approach constant values.

5. Discussion

We found that the smaller time-dependent Lyapunov numbers change even after enough long time in the mass sheet models. This is one of the characters which are expected in the stable chaos. If we find that the time scale of transition from the stable to the ergodic chaos is distributed according to a power law, we can conclude confidently that in this system stable chaos appears. We will analyse this distribution in the near future. Moreover we will analyse systems with much larger numbers of sheets. We might be able to recognise clearly the changing of all the time-dependent Lyapunov numbers and the distribution functions themselves. Furthermore we will analyse whether the stable chaos appears in the real 3-dimensional systems.

References

Aizawa Y., et al., 1989, Prog. Theor. Phys. Supplement, 98, 36
Davies R.L., et al., 1983, ApJ, 266, 41
Goldhirsch I., Sulem P-L., Orszag S.A., 1987, Physica, 27D, 311
Konishi T., Kaneko K., 1992, J.Phys. A, Math. Gen., 25, 6283
Milani A., Nobili A.M., 1992, Nature, 357, 569
Sakagami M., Gouda N., 1991, MNRAS, 249, 241
Yamashiro T., Gouda N., Sakagami M., 1992, Prog. Theor. Phys., 88, 269

Stability of the Modified Konishi-Kaneko System

Shogo INAGAKI

Department of Astronomy, Faculty of Science, Kyoto University,
Sakyo-ku, Kyoto 606-01, Japan

Abstract. The Konishi-Kaneko system is modified to a continuous time system. The dynamical and thermodynamical stability is investigated in this paper. The dynamical stability is investigated with the use of the collisionless Boltzmann equation. The predicted growth rates agree excellently with those obtained from N-body simulations.

Thermodynamic arguments show that the system has clustered equilibrium states as well as uniform states for $1/T > 4\pi/kn$, where k is a coupling constant and n is the number density of particles. The clustered equilibrium is thermodynamically stable and the uniform state is unstable for $1/T > 4\pi/kn$. This stability criterion is the same as that of the dynamical stability.

1. Introduction

Self-gravitating systems have several difficulties to treat with. 1) The potential and force both diverge at zero distance, thus the phase space is non-compact. 2) The range of force extends to infinity, i.e., the force drops very weakly to infinite distance. Though these may be the basic nature of the gravitating systems, they bring about many difficulties to investigate the chaotic nature of the systems. We propose a simpler system but which has some common properties with self-gravitating systems.

Konishi & Kaneko (1992) studied the properties of the N-particle system on a one-dimensional unit circle defined by a symplectic map $(x_i, p_i) \mapsto (x_i', p_i')$:

$$p_i' = p_i + k \sum_{j \neq i}^{N} \sin[2\pi(x_j - x_i)], \qquad k > 0 \tag{1}$$

and

$$x_i' = x_i + p_i' \qquad (\mathrm{mod}\ 1). \tag{2}$$

They discovered the formation and destruction of clusters. Since the system given by Eqs. (1) and (2) is governed by a universal attractive force, the system may have some analogies to the self-gravitating systems. For example, both systems develop cluster-like order although they are conservative systems. We modify the system given by the map (1) and (2) to a system governed by usual Hamilton's canonical differential equations:

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = k \sum_{j \neq i}^{N} \sin[2\pi(x_j - x_i)] \tag{3}$$

and

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = p_i \qquad (x : \mod 1), \tag{4}$$

which are derived from a Hamiltonian

$$H = \sum_{i=1}^{N} \left\{ \tfrac{1}{2} p_i^2 - \frac{k}{4\pi} \sum_{j \neq i}^{N} \cos[2\pi(x_j - x_i)] \right\}.$$

The system given by Eqs. (3) and (4) has no singularity at $x_i = x_j$ and system size is finite (the length unity). Moreover the system is one-dimensional. Thus the system governed by Eqs. (3) and (4) is much simpler than the usual self-gravitating systems.

It is natural to think that the cluster formation in the Konishi-Kaneko system is similar to the Jeans instability in self-gravitating systems. So we investigate the dynamical instability of the system governed by Eqs. (3) and (4) with the use of the collisionless Boltzmann equation. We found that the growth rates of instability derived from the collisionless Boltzmann equation agree excellently with those obtained by N-body simulations.

We next consider the thermodynamic stability of the system to examine if the destruction of clusters discovered by Konishi & Kaneko (1992) is due to some thermodynamic instability. Thermodynamic instabilities occur by relaxation processes which are not described by the collisionless Boltzmann equation. We do not need, however, to derive the collision term to investigate the thermodynamic stability if we assume the second law of thermodynamics.

The collisionless Boltzmann equation is presented and the dispersion relation is derived in Sect. 2. The growth rates of the instability are compared with N-body simulations in Sect. 3. In Sect. 4 we consider the thermodynamic equilibrium states. In Sect. 5 we consider the thermodynamic stability.

2. Dispersion Relation Derived from the Collisionless Boltzmann Equation

From Eqs. (3) and (4) we can immediately write down the collisionless Boltzmann equation:

$$\frac{\partial f(x,p,t)}{\partial t} + p\frac{\partial f(x,p,t)}{\partial x} + k\iint \sin[2\pi(x'-x)]\, f(x',p',t)\,\mathrm{d}x'\mathrm{d}p'\,\frac{\partial f(x,p,t)}{\partial p} = 0\,. \tag{5}$$

It is clear that the state of uniform spatial density with an arbitrary velocity distribution, $f = f_0(p)$ is a stationary state. We impose a small perturbation over the stationary state : $f = f_0(p)+\delta f(x,p,t)$. From the linearized equation for δf we obtain the dispersion relation

$$\epsilon(m,\omega) \equiv 1 - \frac{k}{2}(\delta_{m,1} - \delta_{m,-1}) \int \frac{\frac{\partial f_0(p)}{\partial p}}{\omega - 2\pi m p}\,\mathrm{d}p = 0\,. \tag{6}$$

From Eq. (6) we see that there is no collective mode for $m \neq \pm 1$. The dispersion relation (6) has the same form as the gravitational systems (see e.g. Lynden-Bell 1967) but the allowed wave number is only $m = \pm 1$. This explains why the number of the clusters numerically observed in the model (1) and (2) is limited to one.

We adopt the Maxwellian distribution function,

$$f_0(p) = \frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{p^2}{2T}\right) n\,,$$

as the unperturbed state, where n is the number density of the particles and T is the temperature in energy units. From the well-known Nyquist criterion (e.g. Lynden-Bell 1967), we find that the condition for instability is

$$\frac{n}{T} > \frac{4\pi}{k}\,. \tag{7}$$

This is similar to Jeans' stability criterion for gravitational systems. The maximum growth rate occurs at $T = 0$ and is $\omega_{i,\max} = \sqrt{\pi k n}$.

3. Comparison with N-Body Simulations

To examine the validity of the approximation using the collisionless Boltzmann equation we perform numerical integration of Eqs. (3) and (4) using a second order symplectic integrator (Kinoshita et al. 1991). A quiet start procedure (Sellwood 1983) is used to reduce the amplitude of initial perturbations. There exist a linear growing stage over e^{30} in the amplitude so that

the growth rate is quite accurately determined. The growth rates determined by N-body simulations agree with those determined from the collisionless Boltzmann equation at least up to three digits (Inagaki & Konishi 1993).

Thus we find that the cluster formation in the modified Konishi-Kaneko system is described well by the collisionless Boltzmann equation. This result implies that the force is vanishing at short distances and each particle tends to move in the averaged potential.

4. Thermodynamic Equilibrium States

We assume that the system is described by a single-particle distribution function $f(x,p)$ and define the Boltzmann entropy, S, by

$$S = -\iint f(x,p)\,\ln f(x,p)\,\mathrm{d}p\mathrm{d}x\,. \tag{8}$$

We consider the maximum of the entropy under constraints of constant total mass, M, and of constant total energy, E.

From the necessary condition for the maximum of the entropy under the given constraints, $\delta S - \alpha\delta M - \beta\delta E = 0$, we obtain

$$f = A\exp\left[-\beta\left(\tfrac{1}{2}p^2 + \psi\right)\right]\,, \tag{9}$$

where $A = \exp(-1-\alpha)$. Thus we found that the Maxwell-Boltzmann distribution is the thermodynamic equilibrium states.

We find that ψ is written in the form

$$\psi(x) = B\cos(2\pi x) \tag{10}$$

(Inagaki 1993). If $B = 0$, the system has uniform density distribution and $B \neq 0$ gives clustered equilibrium states. It is possible (Inagaki 1993) to show that non-zero solution of B is possible only for

$$\frac{kM}{4\pi T} > 1\,, \tag{11}$$

where $T = 1/\beta$. Inequality (11) shows that there are non-uniform equilibrium states for sufficiently low temperature. The condition (11) is the same as that for dynamical instability of the system discussed above.

5. Stability Analysis

With the use of Eqs. (9) and (10) we can calculate the total energy, E, as a function of the temperature T. The total energy, E, as a function of $1/T$, is called linear series and useful to study the thermodynamic stability of systems (Poincaré 1885; Inagaki & Hachisu 1978; Katz 1978). The energy E is a single-valued function for $T > kM/4\pi$ and there is a branching at $T = kM/4\pi$. It is evident that the system is stable for sufficiently high temperature so that the system is stable for $T > kM/4\pi$. From the theory of linear series, either branch of $T < kM/4\pi$ is stable and the other is unstable.

Next we investigate the stability of the uniform state (the upper branch), using entropy arguments, that is, to investigate whether the entropy of the system is maximum or not. It is sufficient to consider the states with the entropy as high as possible. It is well known such states have Maxwellian velocity distribution (see e.g., Antonov 1962). Thus we assume that

$$f(p, x) = \frac{1}{\sqrt{2\pi T}} \exp\left(-\frac{p^2}{2T}\right) \rho(x) . \tag{12}$$

Then we can consider that the system consists of a gas with the density ρ and the pressure $P = \rho T$.

From the condition that $\delta S - \alpha \delta M - \beta \delta E = 0$, we obtain

$$\frac{1}{T} - \beta = 0 \tag{13}$$

and

$$\frac{\mathrm{d}P}{\mathrm{d}x} + \rho \frac{\mathrm{d}\psi}{\mathrm{d}x} = 0 . \tag{14}$$

Eq. (13) means that T is constant and Eq. (14) is hydrostatic equation. Using the same arguments as in Sect. 4, we see that $\rho \propto \exp(-B \cos x)$ and solutions with $B \neq 0$ exist for $T > Mk/4\pi$.

We now proceed to the second order variations. We consider δT and δx independent variables of variations. We examine the maximum of $\delta^2 S - \beta \delta^2 E$ under the constraint of constant total energy, $\delta E = 0$. To do so, we impose a normalization condition for the perturbation,

$$\delta^2 \Phi \equiv \int_0^1 (\delta x)^2 \rho \, \mathrm{d}x = M . \tag{15}$$

Thus we consider the condition of the maximum of

$$\mathcal{F} \equiv \delta^2 S - \beta \delta^2 - \mu \delta E - \lambda \delta^2 \Phi ,$$

where μ and λ are Lagrange's multipliers.

From the Euler-Lagrange equation for $\mathcal{F}$ we obtain that $\lambda > 0$ if and only if $4\pi < \beta k M$ or

$$\frac{kM}{4\pi T} > 1 \tag{16}$$

(Inagaki 1993). We note that this condition is the same as the condition for dynamical instability, Eq. (7).

Thus all the conditions for dynamical instability, existence of clustered equilibrium, and thermodynamic instability agree.

Acknowledgement. I thank the Yamada Science Foundation to make the trip to Geneva possible.

References

Antonov V.A., 1962, Vestnik Leningrad Univ. 7, 135
Inagaki S., 1993, in preparation
Inagaki S., Hachisu I., 1978, Publ. Astron. Soc. Japan 30, 39
Inagaki S., Konishi T., submitted to Publ. Astron. Soc. Japan
Katz J., 1978, Monthly Notices Roy. Astron. Soc. 183, 765
Kinoshita H., Yoshida H., Nakai H., 1991, Celest. Mech. 50, 59
Konishi T., Kaneko K., 1992, J. Phys. A25, 6283
Lynden-Bell D., 1967, in: Relativity and Astrophysics, vol. 2, J. Ehler (ed.), vol. 9 of Lectures in Applied Mathematics, (Province, Rhode Island: American Mathematical Society), p. 131
Sellwood J.A., 1983, J. Comput. Phys. 50, 337
Poincaré H., 1885, Acta. Math. 7, 259

Mixing Transformations of N Particles Conserving Almost All Classical Integrals

Daniel PFENNIGER

Geneva Observatory, University of Geneva, CH-1290 Sauverny, Switzerland

1. Paradigm for Inhomogeneous Mixing Flows

The Navier-Stokes equation is often used as paradigm for modelling the gas in galaxies. Yet the interstellar medium (ISM) is far from being a smooth common flow. In fact the cold phase of the ISM, containing most of the gas mass, is *essentially* an inhomogeneous medium better described with a fractal model, and where the density can vary by 10 decades over distances much shorter than $< 1\,\mathrm{pc}$ (Pfenniger & Combes 1994). Typically the "mean-free path" of clumps at the $10 - 50\,\mathrm{pc}$ scale, such as molecular clouds, is much larger than their size, and an hydrodynamical description with particles of the size of molecular clouds can be, and has been, tried. However molecular clouds collide and dissolve because they contain internal degrees of freedom due to the smaller clumps moving inside them. For a fractal gas there is no strong reason to choose a particular scale. The resolution required to follow such an inhomogeneous flow at the $100\,\mathrm{pc}$ scale or larger makes the hydrodynamical approach hardly tractable with todays computers.

So for any practical smallest scale one *can* choose (not to be confused with a very small scale that might be retained in the future with very powerful computers), the differentiability of the flow required by the Navier-Stokes equation for modelling the cold ISM cannot be granted. Although the differential fluid equations are in principle inapplicable, those have nevertheless the main virtue to conserve locally some of the classical integrals at the smallest accessible scale. If the physical flow is highly turbulent and fractal at small scale it is illusory to believe that the simulated flow reproduces faithfully the detail of the physical flow. But if the physical flow mixes rapidly a part of its microscopic properties, as in statistical mechanics only the global integrals are conserved at larger scale and at least these quantities are relevant in simulations.

However the finite element Eulerian approach for describing fluids, either by grid schemes (van Albada 1985; Mulder & Liem 1986; Athanassoula 1991) or by beam schemes (e.g. Sanders & Prendergast 1974; van Albada et al. 1982)

ignores the conservation of the local angular momentum, because this quantity is automatically conserved in the fluid differential equation model where the "elements" are infinitesimally small. But with a finite resolution, this is no longer true, small eddies disappear within grid cells; a finite resolution acts typically as a sink of angular momentum. This can perturb significantly simulations of dissipative and turbulent rotating discs; for example the centre of a disc galaxy is then a sink of angular momentum.

Since we don't have a proper paradigm for describing flows such as the cold ISM, we must invent new ones that reproduce an efficient mixing at small scale while conserving the global integrals at larger scales. Several attempts have been made (e.g. Schwarz 1981; Combes & Gerin 1985; Jenkins 1992; Palouš et al. 1993; Binney & Gerhard 1993) along the line of describing the ISM in a Lagrangian way by a set of colliding particles with substantial mean-free paths, as in planetary rings (Brahic 1977). In these works, as well as with the "Smoothed Particle Hydrodynamics" (SPH) approach (Friedli & Benz 1992) collisions are always binary collisions conserving at least mass and momentum. This is computationally inefficient since the algorithms need to detect the nearest neighbours and to treat individual collisions. Furthermore some of these codes (Schwarz 1981, Palouš et al. 1993) do not conserve angular momentum, which is physically hard to justify. In some way, these particle approaches for gas correspond to the stellar dynamical limit in the limit of zero dissipation. In the stellar dynamical case one has a well defined scale separation between the ISM gas and the gas at the stellar scale, and the interaction between the two systems can be neglected for some times.

In this work we consider whether general dissipative transformations on a set of N point mass particles can mimic complex and highly turbulent dissipative systems in a more efficient way than the previous methods. Fast dissipative schemes corresponding to the simultaneous "collisions" of N-particle that conserve exactly almost all the integrals, including angular momentum but not the energy (otherwise the flow would be conservative), are presented.

2. Mixing Transformations with Conservation Laws

We consider a small region of a physical system for a time sufficiently long during which "mixing" physical processes occur rapidly at small scale, but short enough to be able to neglect the changes due to long range interactions which are supposed to occur at scales larger than the considered region. During this time, the global integrals except energy are assumed to be strictly conserved in the subsystem, because the subsystem can be considered for a short time ("infinitesimally short" for the larger scale) as isolated. Therefore we consider the partial problem of describing consistently the global "collisions" of N isolated point mass particles.

2.1 Conservation Laws

The N particles are described in an inertial frame by their positive mass m_i, their positions $\mathbf{x}_i$, and their velocities $\mathbf{v}_i$, $i = 1 \ldots N$. We require that the considered dissipative process conserves the total mass M, center of velocity $\mathbf{V}$ and center of mass $\mathbf{X}$

$$M \equiv \sum_i m_i, \qquad M\mathbf{V} \equiv \sum_i m_i \mathbf{v}_i, \qquad M\mathbf{X} \equiv \left(\sum_i m_i \mathbf{x}_i\right) - M\mathbf{V}t, \quad (1)$$

and the total angular momentum about its center of mass,

$$\mathbf{L} \equiv \sum_i m_i \left(\mathbf{x}_i - \mathbf{X}\right) \wedge \mathbf{v}_i, \quad (2)$$

i.e. the integrals not involving a particular form of interaction. On the other hand the total energy is not supposed to be conserved. This is motivated by the fact that in many physical cases energy is the less well conserved quantity owing to the numerous possibilities of exchanging rapidly energy by radiative processes. In contrast, momentum and angular momentum can practically only be dissipated by mass exchanges (see, e.g., Shu 1992).

2.2 Arbitrary Transformations of Coordinates

First suppose that we apply an *arbitrary* transformation T_0 to the particles coordinates at a fixed time t. We also suppose that the particle masses can be transformed, but the new masses must remain positive. Since the time is not used, we can set $t = 0$ in the following. This yields new masses ($m_i' > 0$), positions ($\mathbf{x}_i'$), and velocities ($\mathbf{v}_i'$)

$$\begin{pmatrix} m_i' \\ \mathbf{x}_i' \\ \mathbf{v}_i' \end{pmatrix} = T_0 \begin{pmatrix} m_i \\ \mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix} . \quad (3)$$

In general, T_0 does not conserve the global integrals, yielding new integrals M', $\mathbf{X}'$, $\mathbf{V}'$ and $\mathbf{L}'$. We suppose that the actual physical transformation is *mixing*, that is, except for the global invariants, the memory of the precise initial conditions are lost. So we should "correct" T_0 in order to restore a process that conserves the global integrals, except the energy. The corrections should be as uniform as possible for all particles, and should not use the detailed information about the old coordinates (the mixing process forgets the precise values of the old coordinates, but not the global integrals except the energy).

2.3 Mass Conservation

For correcting the mass, a simple assumption is to apply a uniform linear mass correction over all the particles:

$$\begin{pmatrix} m_i'' \\ \mathbf{x}_i'' \\ \mathbf{v}_i'' \end{pmatrix} = T_1 \begin{pmatrix} Am_i' + B \\ \mathbf{x}_i' \\ \mathbf{v}_i' \end{pmatrix} , \tag{4}$$

where A and B are constants. Then in order to conserve mass, A and B should satisfy

$$M'' \equiv \sum_i m_i'' = A\Big(\sum_i m_i'\Big) + BN = AM' + BN = M . \tag{5}$$

Clearly there are many possibilities. But in order to guarantee that this correction yields positive masses, we require $A > 0$ and $B > 0$.

For example the pair

$$A = \alpha \frac{M}{M'} , \qquad B = (1-\alpha)\frac{M}{N} , \tag{6}$$

is a possible solution when $0 \leq \alpha \leq 1$ which averages the masses when α is sufficiently small. The most mixing pair is $A = 0$, $B = M/N$ ($\alpha = 0$), and the less mixing pair is $A = M/M'$, $B = 0$ ($\alpha = 1$).

The main point of this subsection is to show that it is always possible to set up transformations conserving the total mass and positive masses.

2.4 Center of Mass and Center of Velocity Conservation

In order to conserve also the center of mass and the center of velocity, we have in general to correct T_1. A correcting transformation is tried in the form of a constant transformation T_2:

$$\begin{pmatrix} m_i''' \\ \mathbf{x}_i''' \\ \mathbf{v}_i''' \end{pmatrix} = T_2 \begin{pmatrix} m_i'' \\ \mathbf{x}_i'' \\ \mathbf{v}_i'' \end{pmatrix} = \begin{pmatrix} m_i'' \\ \mathbf{x}_i'' + \mathbf{C} \\ \mathbf{v}_i'' + \mathbf{D} \end{pmatrix} , \tag{7}$$

where $\mathbf{C}$ and $\mathbf{D}$ are constant vectors. In order to restore the original centers of mass and velocity, we have, with $M''' \equiv \sum_i m_i''' = M'' = M$,

$$\begin{pmatrix} \mathbf{X}''' \\ \mathbf{V}''' \end{pmatrix} \equiv \sum_i \frac{m_i'''}{M'''} \begin{pmatrix} \mathbf{x}_i''' \\ \mathbf{v}_i''' \end{pmatrix} = \sum_i \frac{m_j''}{M''} \begin{pmatrix} \mathbf{x}_i'' \\ \mathbf{v}_i'' \end{pmatrix} + \begin{pmatrix} \mathbf{C} \\ \mathbf{D} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \mathbf{V} \end{pmatrix} , \tag{8}$$

Thus we have $\mathbf{X}''' = \mathbf{X}$ and $\mathbf{V}''' = \mathbf{V}$ when

$$\mathbf{C} = \mathbf{X} - \mathbf{X}'', \qquad \text{and} \qquad \mathbf{D} = \mathbf{V} - \mathbf{V}'' . \tag{9}$$

We have shown here that it is always possible to correct arbitrary transformations in order to conserve also the centers of mass and velocities.

2.5 Angular Momentum Conservation

The transformations T_1 and T_2 provide a new total angular momentum $\mathbf{L}'''$ that is generally different from the original one $\mathbf{L}$. Since constant corrections are insufficient to conserve angular momentum, we adopt the next simplest hypothesis; we suppose a transformation of the velocities only, and having the same form for all particles except for a linear dependence on the own particle position with respect to the center of mass. Although it could be possible to set up more complicated correcting transformations involving both the positions and the velocities, we don't correct the positions on the ground of simplicity. In general, position transformations produce also potential energy changes, and forces modify velocities to first order, and positions to second order.

In order to modify the total angular momentum without changing the center of velocity, the simplest velocity correction must be orthogonal to and linear with the radius vector to the center of mass. Therefore we require

$$\begin{pmatrix} m_i'''' \\ \mathbf{x}_i'''' \\ \mathbf{v}_i'''' \end{pmatrix} = T_3 \begin{pmatrix} m_i''' \\ \mathbf{x}_i''' \\ \mathbf{v}_i''' \end{pmatrix} = \begin{pmatrix} m_i''' \\ \mathbf{x}_i''' \\ \mathbf{v}_i''' + \boldsymbol{\Omega} \wedge (\mathbf{x}_i''' - \mathbf{X}) \end{pmatrix} , \tag{10}$$

where $\boldsymbol{\Omega}$ is a constant vector to specify. This transformation is a solid rotation in velocity space that obviously conserves the mass and the center of mass, but it conserves also the center of velocity, since

$$\begin{aligned} M''''\mathbf{V}'''' &\equiv \sum_i m_i''''\mathbf{v}_i'''' = \sum_i m_i'''\mathbf{v}_i''' + \boldsymbol{\Omega} \wedge \underbrace{\left(\sum_i m_i'''\mathbf{x}_i''' - M\mathbf{X}\right)}_{=0 \text{ by Eqs. (8) and (9)}} \\ &= \sum_i m_i'''\mathbf{v}_i''' = M\mathbf{V} . \end{aligned} \tag{11}$$

By using the identity $\mathbf{a} \wedge (\mathbf{b} \wedge \mathbf{a}) = \mathbf{a}^2\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\,\mathbf{a}$, then the new angular momentum is

$$\begin{aligned} \mathbf{L}'''' &\equiv \sum_i m_i'''' (\mathbf{x}_i'''' - \mathbf{X}) \wedge \mathbf{v}_i'''' \\ &= \sum_i m_i''' (\mathbf{x}_i''' - \mathbf{X}) \wedge \mathbf{v}_i''' + \sum_i m_i''' (\mathbf{x}_i''' - \mathbf{X}) \wedge [\boldsymbol{\Omega} \wedge (\mathbf{x}_i''' - \mathbf{X})] \\ &= \mathbf{L}''' + \sum_i m_i''' (\mathbf{x}_i''' - \mathbf{X})^2 \boldsymbol{\Omega} - \sum_i m_i''' [(\mathbf{x}_i''' - \mathbf{X}) \cdot \boldsymbol{\Omega}] (\mathbf{x}_i''' - \mathbf{X}) . \end{aligned} \tag{12}$$

If we require that the new angular momentum $\mathbf{L}''''$ equals the original one $\mathbf{L}$ we can solve this linear equation for $\boldsymbol{\Omega}$ in term of the coordinates $\mathbf{x}_i'''$ only. This is a familiar equation of solid body dynamics (see e.g. Goldstein 1950).

2.5.1 Moment of Inertia Tensor

First we define the moment of inertia tensor $I(m_i, \mathbf{x}_i, \mathbf{X})$ about the center of mass $\mathbf{X}$, where $\mathbf{x}_i = (x_i, y_i, z_i)$, $\mathbf{X} = (X, Y, Z)$,

$$\begin{array}{ll} I_{xx} = \sum_i m_i (x_i - X)^2 , & I_{xy} = \sum_i m_i (x_i - X)(y_i - Y) = I_{yx} , \\ I_{yy} = \sum_i m_i (y_i - Y)^2 , & I_{yz} = \sum_i m_i (y_i - Y)(z_i - Z) = I_{zy} , \\ I_{zz} = \sum_i m_i (z_i - Z)^2 , & I_{zx} = \sum_i m_i (z_i - Z)(x_i - X) = I_{xz} . \end{array} \tag{13}$$

After some manipulations, we get

$$S\,\boldsymbol{\Omega} = \mathbf{L}'''' - \mathbf{L}''' = \mathbf{L} - \mathbf{L}''' , \tag{14}$$

where S is the symmetric matrix evaluated for $I(m_i'', \mathbf{x}_i''', \mathbf{X})$,

$$S = \begin{pmatrix} I_{yy} + I_{zz} & -I_{xy} & -I_{zx} \\ -I_{xy} & I_{zz} + I_{xx} & -I_{yz} \\ -I_{zx} & -I_{yz} & I_{xx} + I_{yy} \end{pmatrix} . \tag{15}$$

If S is invertible then $\boldsymbol{\Omega} = S^{-1} (\mathbf{L} - \mathbf{L}''')$. In solid body dynamics $-S$ is also called "inertia tensor" (Golstein 1950), not to be confused with I.

2.5.2 Matrix Inverse

Explicitly, if we write

$$S_x \equiv I_{yy} + I_{zz} , \quad S_y \equiv I_{zz} + I_{xx} , \quad S_z \equiv I_{xx} + I_{yy} , \tag{16}$$

then the determinant $|S|$ reads

$$|S| = S_x S_y S_z - 2 I_{xy} I_{yz} I_{zx} - S_x I_{yz}^2 - S_y I_{zx}^2 - S_z I_{xy}^2 , \tag{17}$$

and the inverse matrix S^{-1} has the form

$$S^{-1} = \frac{1}{|S|} \begin{pmatrix} S_y S_z - I_{yz}^2 & I_{yz} I_{zx} + I_{xy} S_z & I_{xy} I_{yz} + I_{zx} S_y \\ I_{yz} I_{zx} + I_{xy} S_z & S_z S_x - I_{zx}^2 & I_{zx} I_{xy} + I_{yz} S_x \\ I_{xy} I_{yz} + I_{zx} S_y & I_{zx} I_{xy} + I_{yz} S_x & S_x S_y - I_{xy}^2 \end{pmatrix} . \tag{18}$$

2.5.3 Degenerate Cases

If S is not invertible, because $|S| = 0$, we have an infinite number of possible solutions to Eq. (14). In such cases, as a proxy of S^{-1} it is recommended to take the Moore-Penrose pseudo-inverse $S^{\dagger}$ (Moore 1920; Penrose 1955), yielding the unique solution with optimal properties in the least-squares sense.

The pseudo-inverse $A^{\dagger}$ of any matrix A is the unique matrix that satisfies $A^{\dagger} A A^{\dagger} = A^{\dagger}$, $A A^{\dagger} A = A$, and both $A^{\dagger} A$ and $A A^{\dagger}$ are symmetric. The derivation of the pseudo-inverse can be found symbolically (most easily by computer algebra) with the relation

$$A^{\dagger} = \lim_{\varepsilon \to 0} \left[(A^T A + \varepsilon^2 \mathrm{Id})^{-1} A^T \right] , \tag{19}$$

where A^T is the transpose of A and Id is the identity matrix (Frawley 1985).

By a rotation and a translation the moment of inertia tensor I can be made diagonal, and then all the crossed moments I_{xy}, I_{yz}, and I_{zx} vanish.

In that case S is also diagonal and is not invertible only in the case where at least two of the three non-negative tensor components I_{xx}, I_{yy} and I_{zz} vanish (see Eq. (15)). This means that all the points are colinear, i.e. have the form $\mathbf{x}_i = \mathbf{a}t_i + \mathbf{b}$, where $\mathbf{a} \equiv \{a_x, a_y, a_z\}$ and $\mathbf{b}$ are constant vectors.

In this case, S takes the form

$$S = \frac{K}{|\mathbf{a}|^2} \begin{pmatrix} a_y^2 + a_z^2 & -a_x a_y & -a_z a_x \\ -a_x a_y & a_z^2 + a_x^2 & -a_y a_z \\ -a_z a_x & -a_y a_z & a_x^2 + a_z^2 \end{pmatrix} \tag{20}$$

where

$$K \equiv I_{xx} + I_{yy} + I_{zz} = \frac{1}{M} \sum_i \sum_{j>i} m_i m_j (t_i - t_j)^2 \,. \tag{21}$$

If $K > 0$ then the use of Eq. (19) leads to the pseudo-inverse $S^\dagger$

$$S^\dagger = K^{-2} S \,. \tag{22}$$

If $K = 0$ then all the points are concomitant and $S^\dagger = 0$.

2.5.4 Case $N = 2$

In this particular case S is always not invertible since two points are always colinear. Explicitly the pseudo-inverse $S^\dagger$ is

$$S^\dagger = \frac{m_1 + m_2}{m_1 m_2} \frac{1}{(\mathbf{x}_1 - \mathbf{x}_2)^2} \times$$
$$\times \begin{pmatrix} (y_1-y_2)^2+(z_1-z_2)^2 & -(x_1-x_2)(y_1-y_2) & -(z_1-z_2)(x_1-x_2) \\ -(x_1-x_2)(y_1-y_2) & (z_1-z_2)^2+(x_1-x_2)^2 & -(y_1-y_2)(z_1-z_2) \\ -(z_1-z_2)(x_1-x_2) & -(y_1-y_2)(z_1-z_2) & (x_1-x_2)^2+(y_1-y_2)^2 \end{pmatrix} . \tag{23}$$

Of course, if $\mathbf{x}_1 = \mathbf{x}_2$, we have further degeneracy, and then $S^\dagger = 0$.

3. Examples: Linear Friction Laws

3.1 Binary Interactions

Often dissipative particle schemes (Brahic 1977, Schwarz 1981) use binary interactions. In this case one can consider successively pairs of particles, so we need to consider a dissipative law for $N = 2$. As a first guess a linear friction law with respect to the center of velocity is attempted

$$T_0 \begin{pmatrix} m_i \\ \mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix} = \begin{pmatrix} m_i \\ \mathbf{x}_i \\ \mathbf{v}_i - \alpha(\mathbf{v}_i - \mathbf{V}) \end{pmatrix} , \qquad i = 1, 2, \tag{24}$$

where α is a constant. This transformation conserves the total mass and the centers of mass and velocities, therefore T_1 and T_2 are identity transformations. But this friction law does not conserve in general angular momentum. Indeed the angular momentum difference is then

$$\mathbf{L}''' - \mathbf{L} = \mathbf{L}'' - \mathbf{L} = -\alpha \frac{m_1 m_2}{M} (\mathbf{x}_1 - \mathbf{x}_2) \wedge (\mathbf{v}_1 - \mathbf{v}_2) \, .$$

Calculating explicitly the transformation T_3 by using the pseudo-inverse $S^\dagger$ in Eq. (23), we get, after algebraic manipulations,

$$T_3 \begin{pmatrix} m_i \\ \mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix} = \begin{pmatrix} m_i \\ \mathbf{x}_i \\ \mathbf{v}_i - \alpha \frac{m_j}{M} \left[\mathbf{d}_{ij} \cdot (\mathbf{v}_i - \mathbf{v}_j) \right] \mathbf{d}_{ij} \end{pmatrix}, \qquad i = 1, 2, \tag{25}$$

where $j = 3 - i$ ($j = 2$ if $i = 1$ and vice versa), and

$$d_{ij} = \frac{\mathbf{x}_i - \mathbf{x}_j}{|\mathbf{x}_i - \mathbf{x}_j|} . \tag{26}$$

This is precisely the friction law used by Brahic (1977). Only the relative velocity component parallel to the interparticle vector is reduced. This transformation conserves all the integrals but the energy. Noting

$$C_1 \equiv \frac{m_1 m_2}{2M} \left[\mathbf{d}_{12} \cdot (\mathbf{v}_1 - \mathbf{v}_2) \right]^2 \geq 0 \, , \tag{27}$$

the final kinetic energy difference is then

$$\Delta E \equiv \tfrac{1}{2} \left[m_1 (\mathbf{v}_1'''^2 - \mathbf{v}_1^2) + m_2 (\mathbf{v}_2'''^2 - \mathbf{v}_2^2) \right] = -\alpha (2 - \alpha) C_1 \, . \tag{28}$$

So if $0 < \alpha < 2$ the kinetic energy change is negative. The kinetic energy change vanishes when either the vectors of position or velocity difference, $\mathbf{x}_1 - \mathbf{x}_2$ or $\mathbf{v}_1 - \mathbf{v}_2$, are zero, or when these vectors are orthogonal.

In order to find α for a specified energy change ΔE, we solve Eq. (28) for α, and find,

$$\alpha = 1 \pm \sqrt{1 + \frac{\Delta E}{C_1}} \, . \tag{29}$$

This provides the two possible transformations

$$\mathbf{v}_i'''' = \mathbf{v}_i - \left\{ 1 \pm \sqrt{1 + \frac{\Delta E}{C_1}} \right\} \frac{m_j}{M} \left[\mathbf{d}_{12} \cdot (\mathbf{v}_1 - \mathbf{v}_2) \right] \mathbf{d}_{ij} \, . \tag{30}$$

The $(-)$solution is the one which is continuous with no change in velocity when $\Delta E = 0$. The $(+)$solution reverses the sense of the relative velocity.

When the energy change is sufficiently small, $|\Delta E| \ll C_1$, by expanding the square root the $(-)$solution is approximately

$$\mathbf{v}_i'''' \approx \mathbf{v}_i + \frac{\Delta E}{m_i \mathbf{d}_{12} \cdot (\mathbf{v}_1 - \mathbf{v}_2)} \mathbf{d}_{ij} \,. \tag{31}$$

When the energy change is large and positive, $\Delta E \gg C_1$, the approximate solution reads

$$\mathbf{v}_i'''' \approx \mathbf{v}_i \mp \sqrt{\frac{2\Delta E}{M} \frac{m_j}{m_i}} \, \mathbf{d}_{ij} \,. \tag{32}$$

The minimum energy change compatible with this type of dissipation is given by

$$\Delta E \geq -C_1 \,. \tag{33}$$

3.2 N Interacting Particles

Instead of considering particle interactions by pairs, it is much more efficient to group the interactions with larger N. This is justified if the mixing process has time to mix the properties of the N particles during a finite time-step. If a linear friction law is used for an arbitrary number N of particles, we would have

$$\begin{pmatrix} m_i'' \\ \mathbf{x}_i'' \\ \mathbf{v}_i'' \end{pmatrix} = T_0 \begin{pmatrix} m_i \\ \mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix} = \begin{pmatrix} m_i \\ \mathbf{x}_i \\ \mathbf{v}_i - \alpha \mathbf{v}_i \end{pmatrix} , \tag{34}$$

where α is a constant. Then T_1 is an identity transformation and

$$\begin{pmatrix} m_i''' \\ \mathbf{x}_i''' \\ \mathbf{v}_i''' \end{pmatrix} = T_2 \begin{pmatrix} m_i \\ \mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix} = \begin{pmatrix} m_i \\ \mathbf{x}_i \\ \mathbf{v}_i - \alpha(\mathbf{v}_i - \mathbf{V}) \end{pmatrix} . \tag{35}$$

The transformation T_3 is too complicated to give in closed form, but it is straightforward to evaluate numerically Eqs. (13), (15), and (18) efficiently with a computer, since the algorithm is proportional to N. Also in this linear case we do not need to evaluate $\mathbf{L}'''$ since one obtains,

$$S\boldsymbol{\Omega} = \mathbf{L} - \mathbf{L}''' = \alpha \mathbf{L} \,, \qquad \Longrightarrow \qquad \boldsymbol{\Omega} = \alpha S^{-1} \mathbf{L}, \tag{36}$$

where S^{-1} has to be replaced by $S^{\dagger}$ in case of degeneracy. Finally,

$$\mathbf{v}_i'''' = \mathbf{v}_i - \alpha \left(\mathbf{v}_i - \mathbf{V} + S^{-1} \mathbf{L} \wedge (\mathbf{x}_i - \mathbf{X}) \right) . \tag{37}$$

The kinetic energy difference ΔE is given by

$$\Delta E = \tfrac{1}{2} \sum_i m_i \left((\mathbf{v}_i'''' - \mathbf{V})^2 - (\mathbf{v}_i - \mathbf{V})^2 \right) = -2\alpha C_1 + \alpha^2 C_2 \,, \tag{38}$$

where

$$C_1 \equiv \tfrac{1}{2}\sum_i m_i(\mathbf{v}_i - \mathbf{V}) \cdot \left(\mathbf{v}_i - \mathbf{V} + S^{-1}\mathbf{L} \wedge (\mathbf{x}_i - \mathbf{X})\right) ,$$
$$C_2 \equiv \tfrac{1}{2}\sum_i m_i \left(\mathbf{v}_i - \mathbf{V} + S^{-1}\mathbf{L} \wedge (\mathbf{x}_i - \mathbf{X})\right)^2 \geq 0 . \tag{39}$$

Solving α as a function of ΔE, we obtain

$$\alpha = \frac{C_1}{C_2} \pm \sqrt{\left(\frac{C_1}{C_2}\right)^2 + \frac{\Delta E}{C_2}} , \tag{40}$$

where the $(-)$solution is the one continuous with the identity at small dissipation ($\alpha \to 0$ when $\Delta E \to 0$).

The most negative energy change is bounded by

$$\Delta E \geq -\frac{C_1^2}{C_2} . \tag{41}$$

For the $(-)$solution, when dissipation is weak ($|\Delta E| \ll C_1^2/C_2$), we have small energy changes and we get

$$\alpha \approx -\tfrac{1}{2}\Delta E/C_1 . \tag{42}$$

For large positive energy changes ($\Delta E \gg C_1^2/C_2$) we get

$$\alpha \approx -\sqrt{\Delta E/C_2} . \tag{43}$$

Finally when $\alpha = 2$ we have the particular case of strong mixing and yet energy conservation, as in isothermal processes.

4. Conclusions

We have shown how to treat consistently a set of rapidly mixing particles. It is possible to devise various dissipative transformations and then correct them to conserve the mass, centers of mass and velocity, and especially the total angular momentum. Furthermore, for simple linear friction laws it is possible to specify in advance the energy dissipation. In that case the conservation of the other integrals constraints the energy loss not to exceed a maximum value.

Among all the possible corrective transformations that we have described above, only a small part may be applicable in actual physical systems. In real systems further constraints restrict the possible mixing and thus the dissipation rate, e.g., when a flow expands its internal viscosity is much smaller than when it contracts. Only studies with specific problems will allow us to choose the "right" way to model dissipation and mixing. At least the formulation here is general and consistent with the most fundamental constraints that fluids are thought to follow. For example in a shock, another instance where

the hydrodynamic description fails, the shock conditions require precisely to conserve mass and momentum. Here not only these quantities, but also the angular momentum are automatically conserved.

Presently we are experimenting with these transformations. We have implemented the correcting transformations in simple N-body systems ($N \lesssim 100$, and in large-scale N-body simulations ($N \approx 2 \cdot 10^5$). It is straightforward to check that the integrals are indeed conserved exactly. Applications to disc galaxies are investigated in which self-gravity is calculated with a Particle-Mesh method in a 3D polar grid (Pfenniger & Friedli 1991, 1993). This approach allows to simulate gas at large scale much more efficiently that traditional hydrodynamical codes, and is not less realistic with respect to the ISM than the hydrodynamical models since the ISM is highly inhomogeneous at small scale, and no global equation of state is known. In comparison with the SPH technique, 10 times more particles runs 10 times faster with the law discussed in Sect. 3.2, essentially because there is no need to find the nearest neighbours. Instead all the particles within the same grid cells used for the gravitation calculation mix their momentum at a specified rate.

References

Athanassoula E., 1991, in Dynamics of Disk Galaxies, ed. B. Sundelius, Chalmers University, Göteborg, p. 71
Binney J., Gerhard O., 1993, in: Back to the Galaxy, S.S. Holt, F. Verter (eds.), AIP, New York, p. 87
Brahic A., 1977, A&A 54, 895
Combes F., Gerin M., 1985, A&A 150, 327
Frawley W.J., 1985, in Applications of Computer Algebra, ed. R. Pavelle, Kluwer, Boston, p. 415
Friedli D., Benz W., 1992, A&A 268, 65
Goldstein H., 1950, Classical Mechanics, Addison-Wesley, Reading
Jenkins A.R., 1992, D.Phil. thesis, University of Oxford
Moore E. H., 1920, Bull. Amer. Math. Soc. 26, 394
Mulder W.A., Liem B.T., 1986, A&A 157, 148
Palouš J., Jungwiert B., Kopecký J., 1993, A&A 274, 189
Penrose R., 1955, Proc. Cambridge Philos. Soc. 51, 406
Pfenniger D., Combes F., 1994, A&A in press
Pfenniger D., Friedli D., 1991, A&A 252, 75
Pfenniger D., Friedli D., 1993, A&A 270, 561
Sanders R.H., Prendergast K.H., 1974, ApJ 188, 489
Shu F.H., 1992, The Physics of Astrophysics, Vol. II, Gas Dynamics, University Science Books, Mill Valley, California
Schwarz M.P., 1981, ApJ 247, 77
van Albada G.D., 1985, A&A 142, 491
van Albada G.D., van Leer B., Roberts W.W., 1982, A&A 108, 76

Symplectic Integration Without Roundoff Error

David J.D. EARN [1 2]

[1] Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK
[2] Physics Department, Weizmann Institute, Rehovot 76100, Israel

1. Introduction

At every moment dynamical systems are being evolved with the aid of numerical integration algorithms. Unfortunately, it is not always easy to determine whether even the qualitative features of differential equations are well-preserved by numerical "solutions". This is especially true for systems that display chaotic behaviour, and particularly so if the apparent chaos is weak and slow as in the evolution of the planetary orbits (see the recent review by Duncan & Quinn 1993).

A numerical integration method is said to be order k if the theoretical or *truncation error* after one time-step Δt is $O(\Delta t^{k+1})$. Truncation error is an unavoidable consequence of taking a finite time-step. Additionally, *roundoff error* occurs if machine arithmetic is not done exactly; this is normally the case if floating-point numbers are used and it introduces unwanted noise to integrations.

With a time-step Δt and floating-point numbers with P significant binary digits we may estimate the relative importance of the two types of error with what we shall call the *error ratio*,

$$\mathcal{R}(P, \Delta t, k) = \frac{2^{-P}}{\Delta t^{k+1}} \,. \tag{1}$$

Strictly, the total roundoff error depends in a machine-specific way on the sequence of arithmetic operations required at each step (2^{-P} is a lower limit for quantities of order unity) and the truncation error is of the form $C\Delta t^{k+1}$ only in the limit $\Delta t \to 0$ (we have ignored the constant C). Nevertheless, Eq. (1) can be used as a rough guide: roundoff may be unimportant if $\mathcal{R} \ll 1$ but dominates the computational error if $\mathcal{R} \gg 1$. To achieve a given level of truncation error, it is much more efficient to use a modest time-step with a high order method than a tiny time-step with a low order method, so in practice $\mathcal{R} \ll 1$ in low order implementations and $\mathcal{R} \gg 1$ in high order implementations (such as those used in most solar system integrations).

Integration errors are usually quantified by monitoring the evolution of constants of the motion, such as the total energy in a Hamiltonian system. For example, the maximum length of solar system integrations has been guided by an energy error tolerance. Quinn & Tremaine (1990) showed that energy errors could be greatly reduced by carefully avoiding (mainly) the bias of computers to round *up* in all floating-point additions. This trick does not eliminate roundoff error but substantially reduces its influence, at a cost of about a factor of 2 in computing speed.

The magnitude of the errors is important, but so is the character of the errors in long-term studies that aim to determine qualitative behaviour. In the case of the solar system, we would like to distinguish regular from weakly chaotic motion so we would like some reassurance that the numerical errors (principally roundoff) do not significantly alter the qualitative features of the orbits.

The simplest way to start in analyzing roundoff error is to consider a system in which the only error is roundoff. This is the case for *maps*. Earn & Tremaine (1991, 1992; hereafter ET) showed that roundoff errors cause artificial drifting across invariant curves in Hamiltonian maps and can even lead to confusion between regularity and chaos. The elimination of roundoff error in area-preserving maps of the plane was first considered by Rannou (1974). ET showed that Rannou's method can be applied to Hamiltonian maps of arbitrary dimension. To eliminate roundoff error, a given map is slightly perturbed so that it maps a lattice of points to itself and can be iterated exactly on a computer. As shown in ET (and Scovel 1991) these *lattice maps* have the same mathematical structure as the original maps, i.e., they are Hamiltonian (unlike the maps induced by applying floating-point arithmetic to the original formulae).

Lattice maps are known to reduce qualitative errors (e.g., ET). Here, the lattice approach is applied to a fourth-order integration algorithm and it is found that *quantitative* errors (in the conservation of integrals) can also be significantly reduced, even in the short term.

2. Symplectic Integration and Lattice Maps

Most numerical integration algorithms are not designed specifically for Hamiltonian systems and do not respect their characteristic properties, which include the preservation of phase space volume with time (Liouville's theorem). This can lead to spurious damping or excitation. Methods that do preserve all the Hamiltonian properties, i.e., for which the time-forward map is symplectic, are called *symplectic integration algorithms* or SIAs (e.g., Channell & Scovel 1990; Sanz-Serna 1992; Yoshida 1993, and references therein).

Since we are mainly interested here in the gravitational N-body problem, we restrict attention to Hamiltonians in potential form, i.e., which can be

written $H = \frac{1}{2}\mathbf{v}^2 + U(\mathbf{x})$. A variety of useful SIAs for such Hamiltonians can be derived from two simple *symplectic shears*,

$$S_{\mathbf{x}}(t)\begin{pmatrix}\mathbf{x}\\ \mathbf{v}\end{pmatrix} = \begin{pmatrix}\mathbf{x}+t\mathbf{v}\\ \mathbf{v}\end{pmatrix}, \qquad S_{\mathbf{v}}(t)\begin{pmatrix}\mathbf{x}\\ \mathbf{v}\end{pmatrix} = \begin{pmatrix}\mathbf{x}\\ \mathbf{v}-t\frac{\partial U}{\partial \mathbf{x}}(\mathbf{x})\end{pmatrix}. \tag{2}$$

For example, the (first order) leapfrog scheme is

$$S_{\mathbf{x}}(\Delta t) \circ S_{\mathbf{v}}(\Delta t)\,. \tag{3}$$

The (second order) time-centered leapfrog scheme is

$$S_{\mathbf{x}}(\tfrac{1}{2}\Delta t) \circ S_{\mathbf{v}}(\Delta t) \circ S_{\mathbf{x}}(\tfrac{1}{2}\Delta t)\,. \tag{4}$$

A fourth order SIA is given by

$$S_{\mathbf{x}}(a\Delta t) \circ S_{\mathbf{v}}(b\Delta t) \circ S_{\mathbf{x}}(c\Delta t) \circ S_{\mathbf{v}}(d\Delta t) \circ S_{\mathbf{x}}(c\Delta t) \circ S_{\mathbf{v}}(b\Delta t) \circ S_{\mathbf{x}}(a\Delta t)\,, \tag{5}$$

where $a = 1/(2(2-\beta))$, $b = 1/(2-\beta)$, $c = (1-\beta)a$, $d = -\beta c$, and $\beta = 2^{1/3}$. This method, which we shall call `Ruth_4`, is the fourth-order member in the class of algorithms introduced by Ruth (1983). (All these schemes are immediately generalizable to separable Hamiltonians $H = K(\mathbf{p}) + U(\mathbf{q})$.) `Ruth_4` was discovered by Ronald Ruth and first published by Forest & Ruth (1990).

Although all integrators derived from Eqs.(2) are symplectic in theory, they are not symplectic if implemented using finite-precision arithmetic. This problem can be overcome by replacing the shears $S_{\mathbf{x}}$ and $S_{\mathbf{v}}$ with *lattice shears*,

$$\tilde{S}_{\mathbf{x}}(t)\begin{pmatrix}m\mathbf{x}\\ m\mathbf{v}\end{pmatrix} = \begin{pmatrix}m\mathbf{x}+[tm\mathbf{v}]\\ m\mathbf{v}\end{pmatrix}, \quad \tilde{S}_{\mathbf{v}}(t)\begin{pmatrix}m\mathbf{x}\\ m\mathbf{v}\end{pmatrix} = \begin{pmatrix}m\mathbf{x}\\ m\mathbf{v}-\left[tm\frac{\partial U}{\partial \mathbf{x}}(\mathbf{x})\right]\end{pmatrix}, \tag{6}$$

where m is a (large) constant integer and $[\cdot]$ denotes the nearest point on an integer lattice in phase space. Particles on lattice points are mapped to lattice points by $\tilde{S}_{\mathbf{x}}$ and $\tilde{S}_{\mathbf{v}}$ since integer additions can be done exactly. As shown in ET, lattice shears are symplectic so lattice leapfrog and lattice `Ruth_4` are *exactly symplectic in practice* despite the use of finite-precision arithmetic. Using a lattice SIA is equivalent to evolving the exact solution of a problem with a Hamiltonian that is slightly different from the original.

The leapfrog methods are not likely to benefit much from the lattice approach for two reasons: (i) For practical time-steps $\mathcal{R} \ll 1$, even if only single-precision (four byte) floating-point arithmetic is used. (ii) To avoid loss of precision in the force and velocity components when using a lattice map we must have $m\Delta t \geq 2^P$ (see Eqs.(6)). Therefore, since computers provide integers within finite limits only, there is always a minimum time-step permissible in a lattice SIA, *independent of the order of the method.* For low order methods such as leapfrog the minimum Δt is typically too large to obtain acceptably small truncation error.

For a fourth-order integrator the role of roundoff error is significant. We concentrate on the `Ruth_4` algorithm here.

3. Orbital Elements

The gravitational two-body problem is fully integrable and is equivalent to the motion of a single particle in a Kepler potential $(-1/r)$. There are five isolating integrals of the motion (e.g., Binney & Tremaine 1987). In celestial mechanics the integrals are normally expressed as geometric quantities known as the *orbital elements* (e.g., Brouwer & Clemence 1961).

The shape of the (elliptical) orbit is determined by its *semi-major axis* a and *eccentricity* e. The current phase of the orbit is given by its *mean anomaly* $\ell(t)$, and $\ell(0) \equiv \ell_0$ is a constant of the motion (not an isolating integral). The orbit's orientation is determined by its *argument of pericentre* ω (the angle in the orbit plane from ascending node to pericentre), *longitude of ascending node* Ω, and inclination I. In terms of the orbital elements, the energy is $E = -GM/2a$ and the magnitude of angular momentum is $h = \sqrt{GMa(1-e^2)}$, where M is the total mass ("sun" plus "planet").

In a perfect integration (no truncation or roundoff error) all the orbital elements would be exactly conserved.

4. Sample Integrations

`Ruth_4` has been tested with the gravitational two-body problem by Kinoshita, Yoshida & Nakai (1991, hereafter KYN). To facilitate comparisons, we use the same initial conditions and time-step as KYN in all the tests reported here, namely $a = 1$, $e = 0.1$, $\ell_0 = \omega = \Omega = I = 0.349\,\text{rad} \simeq 20°$ and $\Delta t = 0.01 \simeq 1/628$ orbital periods $(T = 2\pi\sqrt{a^3/GM})$. Units are specified by $G = 1$ and $M = 1$.

Fig. 1 shows the evolution of the six orbital elements for ten orbital periods of the two-body problem. Integrations were conducted with ordinary `Ruth_4` (dotted curves) and lattice `Ruth_4` (solid curves) each employing eight byte (64 bit) arithmetic (in the top four panels the dotted and solid curves lie on top of one another). With a lattice size of $m = 2^{62}$ the minimum permissible time-step is $2^P/m \simeq 0.002$ so $\Delta t = 0.01$ does not degrade the force or velocity components. Put another way, with $P = 53$ and $\Delta t = 0.01$ we require $m \geq 2^{53}/0.01 \sim 2^{60}$ to prevent loss of precision, so $m = 2^{62}$ is large enough. However, the error ratio is $\mathcal{R}(53, 0.01, 4) \simeq 10^{-6}$ so the lattice approach is not expected to improve the integration noticeably. In Fig. 1 there is no apparent difference in the evolution of the first four elements but there is a very clear reduction of errors in Ω and I in the lattice integration. The explanation is that the `Ruth_4` integrator exactly conserves the angular momentum vector (cf. KYN) so the errors in Ω and I are entirely due to roundoff despite the fact that $\mathcal{R} \ll 1$. The errors in the other elements are dominated by truncation error (on this timescale at least).

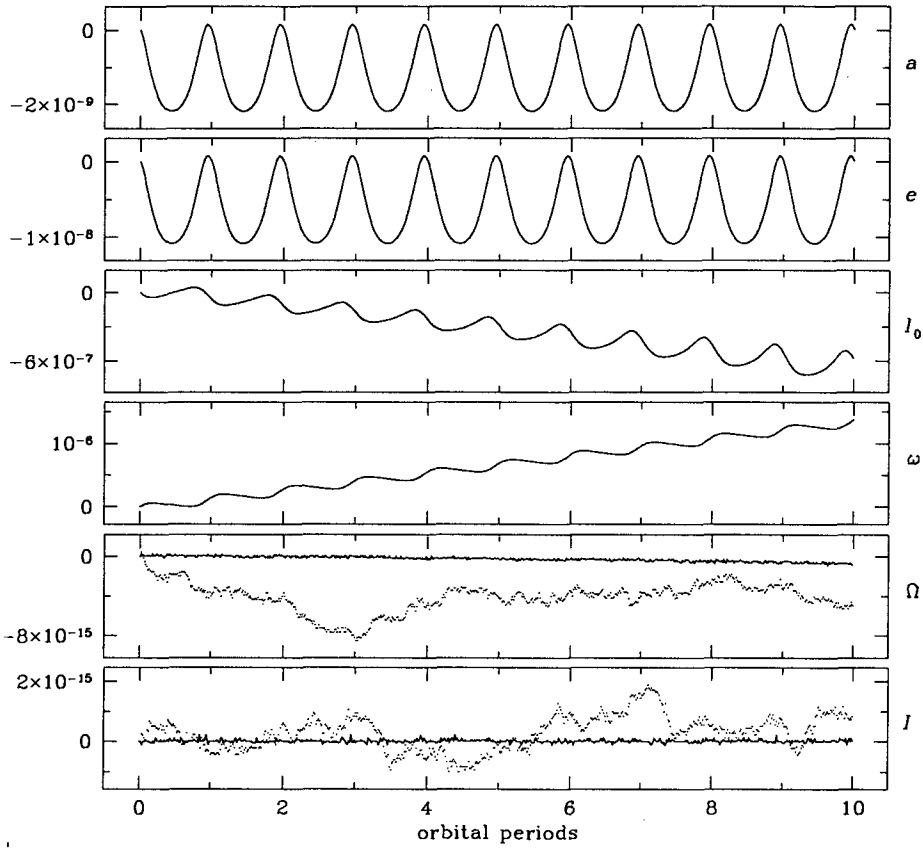


Fig. 1 Errors in the orbital elements during 10 orbital periods of the gravitational two-body problem. Eight byte arithmetic was used with ordinary Ruth_4 (dotted curves) and lattice Ruth_4 with $m = 2^{62}$ (solid curves). The dotted and solid curves coincide in the top four panels. Ω and I (bottom two panels) are more accurately conserved by the lattice method. In the exact solution, all six orbital elements are exactly conserved

In Fig. 1, there are small-amplitude oscillations in a, e, ℓ_0 and ω, the elements dominated by truncation error; such oscillations are expected in any symplectic integrator. In addition, the error in ℓ_0 grows because the numerical scheme has a slightly incorrect period, and the error in ω increases because the numerical potential is slightly non-Keplerian so the orbit precesses. There is no reason to expect a symplectic algorithm to avoid these secular errors. (Of course, they are smaller with higher order SIAs.)

The characteristic oscillations in the elements do not occur if roundoff dominates the computational error. Fig. 2 shows integrations for the same

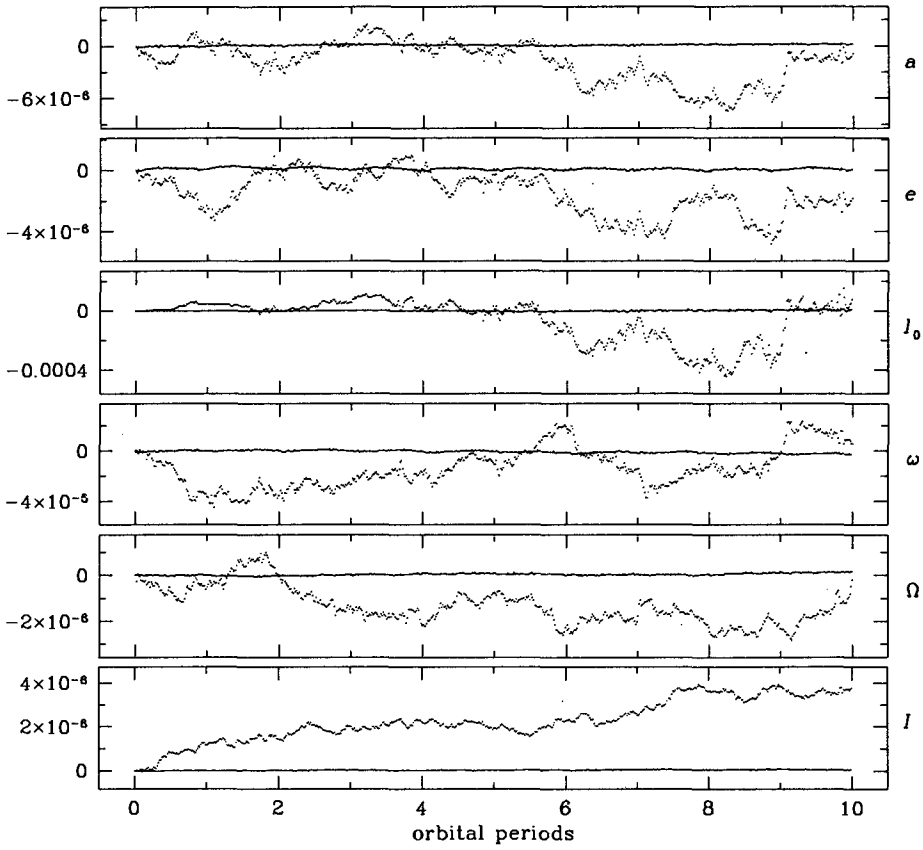


Fig. 2 Integrations of the two-body problem with four byte arithmetic. Lattice Ruth_4 with $m = 2^{30}$ (solid curves) yields more accurate results for all six elements

number of steps as in Fig. 1, this time using four byte (32 bit) arithmetic, i.e., half the precision. The error ratio is now $\mathcal{R}(24, 0.01, 4) \simeq 600$ so the lattice approach is expected to help. However, to avoid degrading the force and velocity components we require $m = 2^{24}/0.01 \sim 2^{31}$. When using four byte integers it is not practical to use $m > 2^{30}$ since this would risk integer overflow from additions, so $m = 2^{30}$ was used for the test shown in Fig. 2. Despite this, lattice Ruth_4 conserves *all* the elements more accurately than ordinary Ruth_4. The improvement is by a factor of about 20 for a, 14 for e, 21 for ℓ_0, 14 for ω, 14 for Ω, and 33 for I (maximum errors in energy and angular momentum are each reduced by factors of about 20). The improvement is greater if the error ratio is larger: repeating the integration with $\Delta t = 0.002$

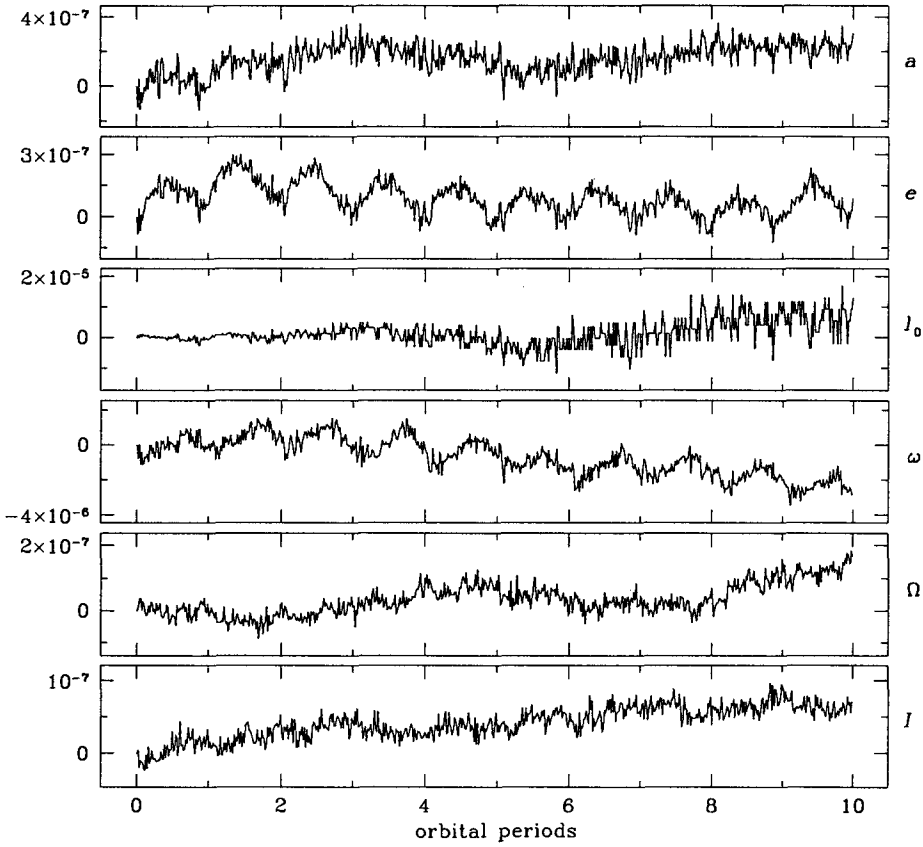


Fig. 3 Expanded scale plots of the results of the lattice Ruth_4 integration shown in Fig. 2. Oscillations typical of SIAs are evident

yields factors of about 36 for a, 51 for e, 34 for ℓ_0, 107 for ω, 22 for Ω, and 42 for I (and 36 for E and h).

Beyond the reduction in the magnitude of the errors that is obvious in Fig. 2, an expanded plot of the lattice integration (Fig. 3) shows that exact symplecticity has restored the expected periodicity in the elements (see e and ω especially). There appears to be an extra oscillation on a timescale of about five orbital periods and in general the curves are nowhere near as smooth as those in Fig. 1. The lack of smoothness is not surprising since to avoid roundoff error the lattice method introduces very small scale fluctuations in the Hamiltonian; these are not present in the Ruth_4 algorithm itself. Note that the periodicity observed in Fig. 3 is *not* the oscillation due to truncation error that is intrinsic to Ruth_4—the amplitude of that is smaller by two orders of magnitude (compare Figs. 1 and 3).

5. Discussion

This paper has shown that the lattice Ruth_4 algorithm outperforms ordinary Ruth_4 when the computational error is dominated by roundoff ($\mathcal{R} \gg 1$). In integrations of the gravitational two-body problem with a time-step $\Delta t = 0.01$, the magnitude of the errors is reduced by a factor of order 20 (Fig. 2) and the character of the errors is more like what is found when truncation error dominates (compare Fig. 3 to the ordinary Ruth_4 integrations in Figs. 1 and 2). The reduction of errors is more significant in going from ordinary Ruth_4 to lattice Ruth_4 than from a fourth-order (non-symplectic) Runge-Kutta integrator to ordinary Ruth_4 (the experiment conducted by KYN).

The time-scale considered here is very short (10 orbital periods) but similar improvement occurs in much longer integrations. Using the lattice approach can clearly have noticeable qualitative effects on the dynamics only after a very long time. Studies that span millions of orbital periods may benefit significantly from the use of lattice SIAs.

Yoshida (1990) extended the Ruth class of SIAs to arbitrary (even) order; the lattice approach applies to all these integrators. More work is necessary to find high order SIAs whose efficiency compares well with the multi-step methods that are currently used in solar system integrations (e.g., Quinn, Tremaine & Duncan 1991). This is a challenging problem because some of the free parameters that could be used to increase order must be used to arrange symplecticity.

To force roundoff to dominate the Ruth_4 integrations, four byte arithmetic was used. The factor by which the elements are conserved better by lattice Ruth_4 increases as the time step is reduced, i.e., as roundoff becomes relatively more important. This suggests that the relative gain achieved by lattice methods is likely to be more significant for higher order integrators (for which $\mathcal{R} \gg 1$ even when using eight byte arithmetic). In long integrations requiring high order methods, such as simulations of the solar system, lattice SIAs may offer the best way to reduce the effects of numerical errors on dynamical evolution.

The $m = 2^{62}$ lattice integration shown in Fig. 1 was done on a Convex computer, which provides full eight byte integers in hardware. Unfortunately, most computers do not supply full length integers in hardware, i.e., the longest integers are not as long as the longest floating-point numbers. Integer arithmetic can be done using the longest floating-point numbers, which typically raises the maximum lattice size from $m = 2^{30}$ to $m = 2^{52}$. This is very valuable for studies of maps (e.g., ET) but it is usually not sufficient to benefit lattice SIAs (because of the requirement that $m\Delta t \geq 2^P$ to avoid degrading the force and velocity).

Lattice methods are useful for studies of the long-term evolution of Hamiltonian systems. For Hamiltonian maps, it is usually advantageous to use the lattice approach. For continuous Hamiltonian systems, lattice methods sig-

nificantly reduce the errors in *high order* SIAs. This improvement is obtained without significantly compromising efficiency *provided* that full length integers are available in hardware. Thus, for a Hamiltonian problem requiring a high order integrator, and a computer with full length integers, a lattice SIA should be used.

Acknowledgments. It is a pleasure to thank Scott Tremaine for helpful advice and comments, and Moti Milgrom for his hospitality at the Weizmann Institute. This research was supported by a Weizmann Institute Exchange Fellowship and a grant from the Albert Einstein Centre for Theoretical Physics.

References

Binney J.J., S. Tremaine, 1987, Galactic Dynamics, Princeton University Press

Brouwer D., Clemence G.M., 1961, Methods of Celestial Mechanics, Academic Press

Channell P.J., C. Scovel, 1990, Symplectic Integration of Hamiltonian Systems, Nonlinearity 3, 231–259

Duncan M.J., T. Quinn, 1993, The long-term dynamical evolution of the solar system, Ann. Rev. Astron. Astrophys.31

Earn D.J.D., S. Tremaine, 1991, Exact solutions for Hamiltonian systems using integer arithmetic, in Dynamics of Disc Galaxies, B. Sundelius (ed.), Göteborg University, Göteborg, Sweden), pp. 137–141

Earn D.J.D., S. Tremaine, 1992, Exact numerical studies of Hamiltonian maps: Iterating without roundoff error, Physica D 56, 1–22

Forest E., R.D. Ruth, 1990, Fourth-order symplectic integration, Physica D 43, 105–117

Quinn T., S. Tremaine, 1990, Roundoff error in long-term planetary orbit integrations, Astron. J.99, 1016–1023

Quinn T., S. Tremaine, M. Duncan, 1991, A three million year integration of the Earth's orbit, Astron. J.101, 2287–2305

Rannou F., 1974, Numerical Study of Discrete Plane Area-preserving Mappings, Astron. Astrophys.31, 289–301

Ruth R.D., 1983, A canonical integration technique, IEEE Trans. Nuc. Sci. 30, 2669–2671

Sanz-Serna J.M., 1992, Symplectic integrators for Hamiltonian problems: An overview, Acta Numerica 1, 243–286

Scovel C., 1991, On Symplectic Lattice Maps, Phys. Lett. A 159, 396–400

Yoshida H., 1990, Construction of higher order symplectic integrators, Phys. Lett. A 150, 262–268

Yoshida H., 1993, Recent progress in the theory and application of symplectic integrators, Celest. Mech. Dyn. Astron.,to appear

Discreteness Noise Versus Force Errors in N-Body Simulations

Junichiro MAKINO

Department of Information Science and Graphics, College of Arts and Sciences, University of Tokyo, Komaba 3-8-1, Meguro-ku, Tokyo 153, Japan

Abstract. We have investigated the effect of the numerical error of the force calculation on the accuracy of the result of collisionless N-body simulations. Our main result is that force calculation with a low accuracy — a few percent for each particle pair — is sufficient, irrespective of the number of particles. Higher accuracy is useless since the discreteness noise, i.e., the two-body relaxation effect, dominates the error. We have performed numerical experiments for two different cases: a special-purpose machine with low-precision hardware and the tree-code. In both cases, the two-body relaxation effect dominates the error and the effect of the error in the force calculation is practically negligible for all cases we tried.

1. Introduction

For a large stellar system, such as a galaxy containing more than 10^{10} stars, the two-body relaxation time is much longer than the age of the universe. These systems can be described by the collisionless Boltzmann equation. To study the time evolution of such systems, we approximate the continuous distribution function by an N-body system with the number of particles smaller than that in real galaxies. The two-body relaxation effect makes the evolution of the N-body system different from that of the original system.

The two-body relaxation is due to the fluctuation in the potential field expressed as the superposition of N point-masses. The potential calculated by the N-body approximation contains the error of the order of $1/\sqrt{N}$. This error in the potential randomly changes the velocity of particles. This random change of the velocity is the two-body relaxation effect itself. In fact, the change of the velocity due to the two-body relaxation per crossing time is $O(\sqrt{N})$, which is consistent with the above description.

The only way to reduce the two-body relaxation effect is to increase the number of particles (Hernquist & Barnes 1990). Therefore, as long as the effect of the numerical error is smaller than the two-body relaxation effect,

one should increase the number of particles at the cost of the numerical accuracy, in order to improve the overall accuracy of the result.

In the paper, we evaluate the effect of the numerical error in the force calculation to the overall accuracy of collisionless N-body simulations. We consider the numerical error introduced by the tree-code (Barnes & Hut 1986) and the low-accuracy calculation on GRAPE-3 hardware (Okumura et al. 1993).

GRAPE-3 is a special-purpose computer designed for the fast time integration of astrophysical N-body systems. In N-body simulations, the calculation of the force between particles consumes most of the CPU time, even when an $O(N \log N)$ algorithms such as the tree-code is used. A GRAPE (an abbreviation for GRAvity PipE) is a specialized hardware for the fast calculation of the gravitational force between particles. It has 48 parallel pipelines, each of which can calculate the gravitational force in the speed of 300 Mflops. Thus the peak speed of GRAPE-3 is 14.4 Gflops. Typical sustained speed is in the range of 5-10 Gflops, depending on the number of particles.

To achieve the high level of parallelism (48 pipelines each of which performs about 30 floating point operations simultaneously), we implemented one pipeline into a custom LSI chip. In order to reduce the physical size (and thus the price) of the chip, it was necessary to reduce the numerical accuracy. We have adopted a combination of the fixed-point and logarithmic number format, so that the relative accuracy of the pairwise force is about 2% and the accumulation of the forces from many particles will not produce no unpredictable round-off errors.

The theoretical basis of the analysis presented here has been described in more details by Makino et al. (1990). Detailed numerical result has been given by Hernquist et al. (1993).

2. Numerical Experiments

We integrated the time evolution of the Plummer model for 3 crossing times using six different methods for force calculation. Table 1 gives the methods we used. In the system of units we used, the half mass radius is 0.77 and the crossing time is $2\sqrt{2}$. We used the standard Plummer softening with the softening length of $\epsilon = 1/32$. Time integration is done using the leap-frog scheme with the time-step of $\Delta t = 1/64$.

We calculated the root-mean-square relative change of the binding energy of par crossing time defined as

$$\sigma_R^2 = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{(E_{i,1} - E_{i,0})}{E_{i,0}} \right]^2 , \tag{1}$$

where $E_{i,0}$ and $E_{i,1}$ are the energies of particle i at the times t_0 and t_1.

Table 1. Methods used

Graphic symbol	Method
Filled circle	N^2 direct summation on a Cray YMP
Open circle	N^2 direct summation on a GRAPE-3
Filled square	Tree-code with $\theta = 0.5$ and quadrupole moment
Cross	Tree-code with $\theta = 1.0$ and quadrupole moment
Filled triangle	Tree-code with $\theta = 0.5$ and monopole moment
Three-armed cross	Tree-code with $\theta = 1.0$ and monopole moment

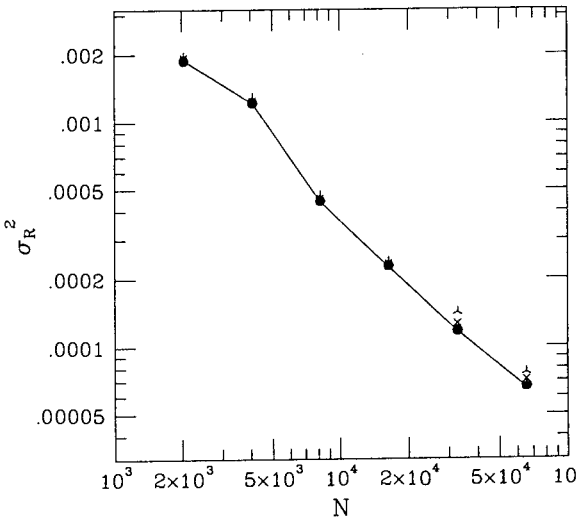


Fig. 1. Rms change in particle energy σ_R for a fixed time interval, as a function of particle number N, for the six methods listed in Table 1

Figure 1 shows σ_R^2 as a function of N, for $2048 \leq N \leq 65536$. The rms energy change is practically the same for all 6 methods used, except that tree-codes with $\theta = 1$ show somewhat larger change for large values of N. In order to separate the energy change due to numerical error from that due to two-body relaxation, we calculated the quantity

$$\delta = \left(\frac{\sigma_R}{\sigma_{R,\mathrm{direct}}}\right)^2 - 1\,, \tag{2}$$

where $\sigma_{R,\mathrm{direct}}$ is the energy change for the runs with the direct summation on a Cray YMP. For tree-codes with $\theta = 1$, δ is around 0.1, for GRAPE-3, $\delta \sim 0.01$, for tree-codes with $\theta = 0.5$, $\delta \sim 0.003$. The dependence of δ on N is very weak. Thus we can conclude that the energy change due to numerical error is smaller than, and proportional to, the two-body relaxation, and therefore would become smaller as we increase the number of particles.

3. Discussion

A simple theory (Makino et al. 1990) predicts that the effect of the numerical error is proportional to the two body relaxation effect (i.e., δ is constant), if the error of the forces between different pairs of particles are independent. Therefore, even if the force calculation contains some constant amount of error, the numerically integrated N-body system converges to the continuous system in the limit of $N \to \infty$, since both the two-body relaxation and numerical error approach zero.

The effect of the numerical error becomes small as we increase the number of particles simply because of cancellation of errors of different forces. If the errors are independent of each other, the numerical error of the force on a particle, which is the summation of N forces, is $O(1/\sqrt{N})$. This dependence is the same as that of the fluctuation of the potential due to the finite number of particles, which is the source of the two-body relaxation effect.

The numerical errors of the tree-code or GRAPE hardware are not really random. Therefore, for a very large N, the energy change due to numerical error would become constant. However, the result of our numerical experiments suggests that for practical values of N the errors can be regarded as random. To summarize, the effect of the numerical error in the force calculation to the accuracy of the collisionless N-body simulation is smaller than the two-body relaxation effect, for all of the six methods we tested, independent of the number of particles.

References

Barnes J.E., Hut P., 1986, Nature 324, 446
Hernquist L., Barnes J.E., 1990, ApJ 349, 562
Hernquist L., Hut P., Makino J., 1993 ApJL 402, 85
Hockney R.W., Eastwood J.W., 1988, Computer Simulation Using Particles, New York, Adam Hilger
Makino J., Ito T., Ebisuzaki T., 1990, PASJ 42, 717
Okumura S. K., Makino J., Ebisuzaki T., Ito T., Fukushige T., Sugimoto D., Hashimoto E., Tomida K., Miyakawa N., 1993, PASJ 45, 349

4. Instabilities

Core Motions and Global Chaotic Oscillations

Richard H. MILLER

Department of Astronomy and Astrophysics, University of Chicago

1. Ancient History

It may be of interest to recall the discovery and early development of the exponential separation of particle trajectories in gravitational N-body problems before we turn to the principal subject of this review.

The exponential separation of trajectories was found as one of those cases in which something unexpected happens in a calculation, and a discovery was made as a result of tracking it down. Call it serendipity. In this case, I foolishly thought that a good way to check new a gravitational N-body program would be to test it for microscopic reversibility. That produced a disaster.

The computer program was patterned after that of von Hoerner (1960). It included a few unusual features, such as the speedup from recomputing force contributions from distant stars infrequently, a speedup later associated with the names of Ahmad & Cohen (1973). I didn't think it was important enough to mention in the paper. Computers were small and slow by today's standards, so experiments were limited to $25-32$ bodies, and many were run with as few as 4, 8, or 12 particles.

Failure to reverse got worse and worse as the experiments were extended to longer times. Runs to T, $2T$, $4T$, and so on were annoying because many of the computer cycles were used re-computing the same interval. There were difficulties associated with reversing the system and re-starting it in the reverse direction as well.

"Parallel calculations", in which two nearly identical systems run forward in time, more or less in lockstep, provided a nice way around this difficulty. They led to the plots of Fig. 1 taken from the original 1964 paper (Miller 1964), in which the logarithm of the trajectory separation in the $6N$-dimensional phase space is plotted as a function of time. The top panel shows an 8-body system, while a 12-body system is in the bottom panel. The underlying linear trend in these plots indicated exponential separation.

The rapidity of the separation and the huge increase above the original separation both struck me at the time. In a very few crossing times, trajectory

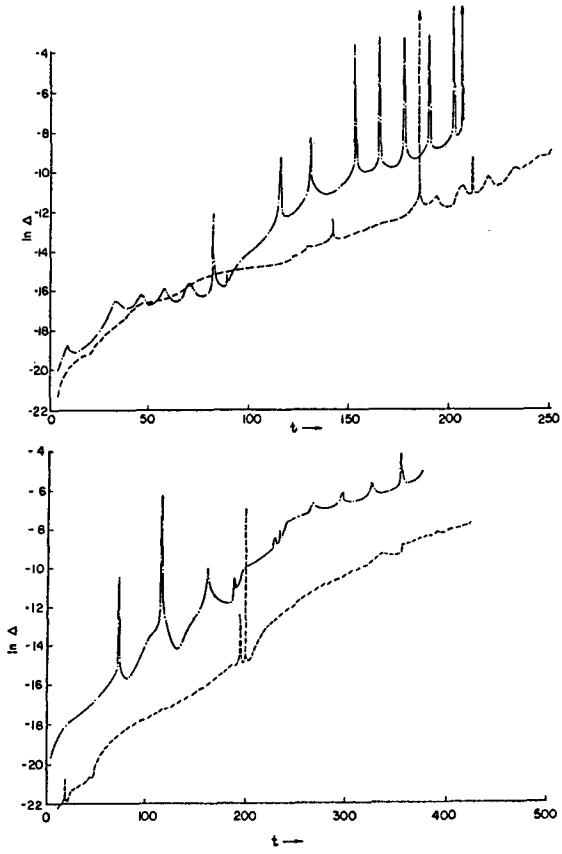


Fig. 1.
Trajectory separations

separation could become large enough that the two systems would no longer realize they had a common parentage. And this started with *very* small initial separations. N-body systems with different potentials, such as the molecular 6 – 12 potential, acted differently. There was still exponentiation, but the continued slope between close encounters was absent.

Standish (1968) soon confirmed the effect and made the first study of the effect of softening near encounters. At about that same time, Lecar undertook to define a "standard" initial condition for gravitational N-body calculations that would allow workers to verify their codes, although he was warned of impending disaster. Results from that check provide the one of the most dramatic demonstrations of the power of the exponentiation that has ever been published (Lecar 1968).

I didn't know the language of Lyapunov characteristic numbers or of Kolmogorov entropy at the time. Some time later, I found that Krylov (1979) had essentially described the process underlying the separation. Recall that all this was long before chaos theory became fashionable—it came at about

the same time as the famous paper describing the Lorenz Attractor (Lorenz 1963).

Although there is an exponential process, and the characteristic time scale of any exponential process can be called a "relaxation time", the e-folding time associated with trajectory separation is *not* the same as the standard relaxation time of stellar dynamics. A system may have several relaxation times. Think of the spin-echo system. This is simply another relaxation time within a stellar dynamical system in addition to 2-body and mean field (violent) relaxation. Exponentiation and 2-body relaxation have been mistakenly associated in the past.

Several questions from that time are still open.

1. Estimate the e-folding time from first principles,
2. The N-dependence of the e-folding rate, and
3. How can we draw valid physical inferences from a calculation in which the underlying physical process is chaotic?

You've heard some recent work on (1) and (2) at this Workshop. Simple analytic estimates for (1) gave $1/\sqrt{G\rho}$, but they usually contain some untenable assumptions (principally uncertain estimates of the importance of pairing terms). An early attempt is in Miller (1966).

One attack on question (3) led to a time-reversible integrator (Miller & Prendergast 1968), that would today be called symplectic. It gives a bit of confidence in integration methods, but it does not provide a completely satisfactory answer. Its main use turned out to be in problems with large N, such as in the studies of galactic dynamics to be reported later in this talk. Question (3) was put to numerical analysts (Miller 1974), but I know of no answers apart from suggestions to average over an ensemble of experiments (Miller 1967; Smith 1979).

I've run a few additional checks since then, some of which have been published (see, e.g., Miller 1971a, 1971b), others not. A study of the dimensionality of the locally expanding and contracting subspaces along the actual trajectory of an N-body system was the most interesting of the unpublished studies.

Perturbation equations for the growth of separation take the form

$$\dot{\xi} = M\xi \,, \tag{1}$$

where ξ is a $6N$-vector of differences of momenta and of coordinates, and M is a $6N \times 6N$ matrix in the space of the ξ's (Miller 1971a, 1974). If the unperturbed problem is written in cartesian coordinates, the matrix M is,

$$M = \begin{pmatrix} 0 & \frac{1}{m}I \\ (\text{grad } F) & 0 \end{pmatrix} \,, \tag{2}$$

where 0 and I denote $3N \times 3N$ block matrices of all zeroes or the identity, m is the mass of a particle (here taken all equal), and (grad F) denotes the

$3N \times 3N$ block matrix of force gradients. It is built up of 3×3 blocks, and each of those little blocks represents the tensor force gradient between two particles within the total configuration. The total matrix, (grad F), is real symmetric, and its trace is zero. It is clear that a ($3N$-dimensional) rotation of the configuration space that brings (grad F) to diagonal form also rotates the momentum space, but the identity section of M remains diagonal. Thus the diagonal form (the eigenvalues) of (grad F) gives a decomposition of the configuration space in the neighborhood of the actual particle configuration. The eigenvalues give instantaneous e-folding rates: they are akin to Lyapunov numbers. There are $3N$ of them and they sum to zero. The positive eigenvalues span a subspace that is locally expanding, and the negative eigenvalues span another (orthogonal) subspace that is locally contracting. The number of dimensions of each of these subspaces is of interest. Because the individual 3×3 blocks have an eigenvalue pattern of $(+2, -1, -1)$, we expect about twice as many negative eigenvalues as positive.

The experiments mentioned consisted of studying the eigenvalues of (grad F) as an N-body calculation proceeded. The number of positive and negative eigenvalues changes as the integration proceeds, always keeping about twice as many negative eigenvalues as positive. But the change is the important feature. An Anosov system retains the *same* number of dimensions in the expanding and contracting (and neutral) subspaces throughout the available configuration space. The gravitational N-body system has thus been demonstrated *not* to be an Anosov system. (Anosov systems are called C-systems by Arnold & Avez 1968.)

A second test was to check the e-folding rate in a 60,000-body problem. It showed a growth rate of a few e-foldings per crossing time. But it was run with a grid code and softened particle interactions, and so is of limited value for the general problem of determining how e-folding rates depend on particle number.

2. Core Motions

A galaxy's nucleus orbits around its mass centroid. Orbital motions are overstable in numerical experiments started with the nucleus at rest atop its mass centroid. The amplitude doubles in $6-10$ orbital periods. Orbits precess, nutate, and change their amplitudes, but they keep fairly constant periods. Amplitudes reach a core radius. Orbital periods are in resonance with local particle motions, indicating that center motions are a local, rather than a global, phenomenon. The overstability implies that a galaxy cannot be formed in Nature with its nucleus at rest atop its mass centroid, and that nuclei orbit the mass centroid in real galaxies. A fuller report of this work has been published elsewhere (Miller & Smith 1988, 1992), and these references should be consulted for additional details. A videotape showing dynamical developments accompanies the 1992 reference.

Motions of galactic nuclei with respect to their mass centroids were discovered in numerical experiments with a disk embedded in an axisymmetric galaxy. The evolution of a disk inclined relative to the symmetry axis of an oblate galaxy was being studied with the goal of determining the long-time integrity of the disk. Center motions became apparent on a new graphics display. Disk particles near the center were moving about in an unexpected manner. They move as a more or less coherent batch; a little patch of the innermost disk particles remains reasonably planar throughout the experiment. The normal to that plane gyrates smoothly with epicyclic motion and nutation superposed atop a basic precession. The motions are very complex. Epicyclic motions and nutation might have been expected, but the patch builds up an orbital motion as well, and that orbital motion was not expected.

A particle initially at rest at the center of a galaxy starts to oscillate, and its amplitude continues to grow thereafter. That particle is not alone. It is part of a density wave that builds up as many particles join a group that moves together in the local (harmonic) potential of the galaxy. Other particles move with it, and motions of the collective mass cause (time-varying) distortions to the (local) gravitational potential down near the galaxy's center. These distortions sweep up more stars in the neighborhood and, in turn, the added mass of those additional stars generates a larger distortion of the potential. Distortions of the potential field are amplified as more and more mass gets swept up in the motion. That drives overstable motions. The amplitude doubles in about $6 - 10$ cycles of the basic oscillation. In three dimensions, the orbital amplitude grows and the center follows a rather complicated trajectory. These center motions are a physical effect, and their growth indicates an overstability (or instability).

The match of periods of center motions with orbital periods suggests that center motions are resonant and are a local phenomenon. They are not global motions in which the whole galaxy rings in a normal mode. The growth terminates when amplitudes get large enough to explore regions where the basic potential of the underlying galaxy is no longer harmonic.

Galaxies whose nuclei are off-center with respect to neighboring isophotes provide observational evidence for center motions. Nearly every galaxy that has been studied carefully enough to show the effect has a nucleus that is off center, usually by a few arcseconds. Several examples are discussed in Miller & Smith (1992), and we have found a few more since that paper was published. Again, the reader is referred to that paper for more detail.

The overstability (instability) implies that a galaxy could not exist in Nature with its nucleus at rest atop its mass centroid. It could never have been formed that way in the first place because that condition is unstable. High sensitivity (reasonable signal-to-noise ratio) made it possible to discover that the standard initial condition for numerical experiments, with the nucleus at rest at the mass centroid, is unstable.

Galaxies in experiments do, however, develop into stable systems that are capable of representing real galaxies. Their stability is demonstrated by the fact that they survive for reasonably long times. Fortunately, they are not terribly different from the starting models—they differ by the orbiting nucleus. The process by which galaxies form can hardly be gentle, so it is difficult to imagine that a real galaxy in Nature could reach a quiet state in which its nucleus remains centered.

Central regions of galaxies and of star clusters are almost certainly not in a static steady state. Most observational inferences about the nuclear regions of a galaxy (amount and distribution of mass, velocity fields, and so on) are tacitly based on the idea that the region is in a static steady state. Many of these conclusions are shaky if the nuclear regions are sloshing about.

3. Global Chaotic Oscillations

The center motions described in the previous section are local disturbances, but they demonstrate that the global smoothed gravitational potential changes with time. It is not static. We next question whether there are global disturbances as well. It turns out that there are, and that they can have surprisingly large amplitudes. This means that galactic potentials show even more important temporal variations.

3.1 Oscillations in the Total Kinetic Energy

Pulsations in the total kinetic energy provide the most direct manifestation of galactic oscillations. Large amplitudes are evident in Fig. 2, which shows the time dependence of the total kinetic energy within a spherical galaxy model, our experiment 13. The peak-to-peak variation in kinetic energy is 22% of the mean; rms variations are 8%. A crossing time $T_{cr} = 25$ of the units on the abscissa. The kinetic energy oscillates with period 83. No decrease in amplitude can be seen in the plot, over some 15 full cycles of the oscillation. This is about $51T_{cr}$ (5 billion years for a typical galaxy); the experiment was later extended for another $10T_{cr}$, still with no change in oscillation amplitude. The upper track shows the total potential energy with its sign reversed.

Total energy is conserved to within 0.05% over the duration of this experiment (and of all the experiments of this sequence). Total potential energy and total kinetic energy change together to leave the sum conserved.

The plot in Fig. 2 is remarkable. Wiggles in this curve would be expected to damp out to a constant value (due to phase mixing or to violent relaxation) within a few cycles according to conventional theory, but these oscillations continue for long times. Their amplitude is too large and their period is too stable for these wiggles to represent random fluctuations of the kinetic energy. This result is consistent with the virial theorem, which does not rule

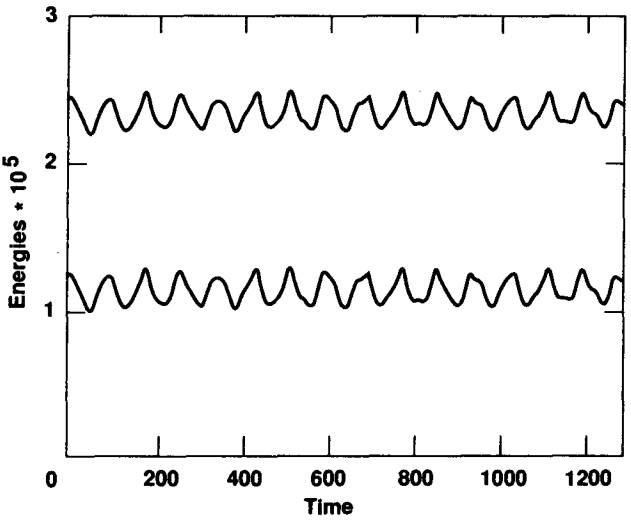


Fig. 2. Total kinetic energy (lower track) and potential energy (upper track; sign reversed) vs. time for experiment 13

out periodic variations with large amplitude. It only makes statements about time averages.

Further, there are some peculiar symmetries, such as the reflection symmetries around times 350, 800, and 980. Some characteristics of this plot suggest chaos, as do similar plots with variable stars (Perdang 1985).

Plots of total kinetic energy for additional experiments of the same sequence show continuing oscillations with equally large amplitude. The time variation is more nearly sinusoidal, lacking the peculiar features suggestive of chaos in Fig. 2.

The envelope of oscillations from these experiments is quite uniform. No damping can be seen over $50 - 60$ crossing times. Damping by 5% could be seen; we can set an upper limit on the order of 0.1% per crossing time for either experiment, whatever the cause of damping. Landau damping is one possible cause, and mode-mode coupling is another. If present, the effect of either must be very weak. The oscillation amplitude is independent of particle number. It is still pretty clean, and it continues to show no appreciable damping, with as few as 10 000 particles. Most experiments in the sequence were run with 400 000 particles.

This plot clearly demonstrates persistent oscillations with surprisingly large amplitudes and with no detectable change in amplitude over the duration of the experiment. The important result of this investigation is evident from this plot alone. Experiment 13 represents a breakthrough because of the large amplitude. Other properties stand out clearly above the noise as well, so it is a good case to study.

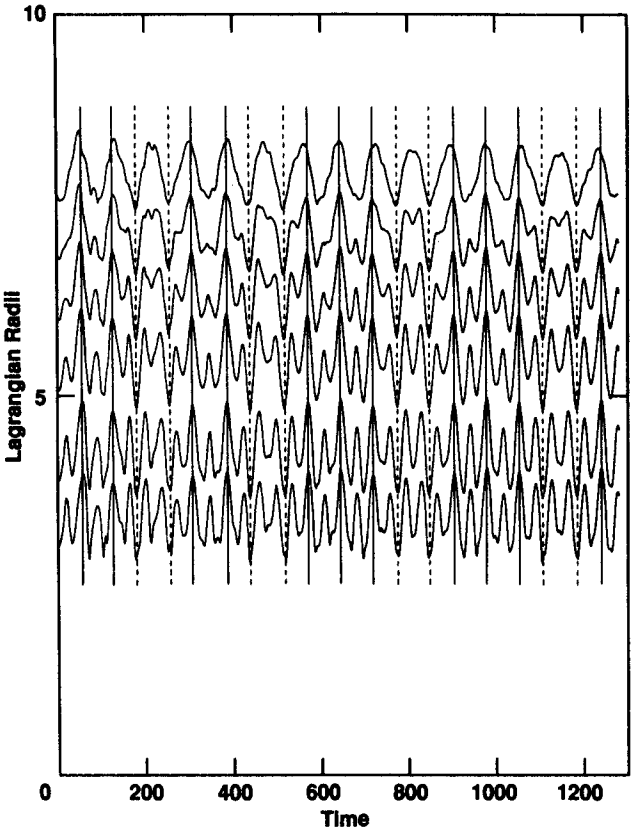


Fig. 3. Lagrangian radii vs. time for experiment 13

We turn next to additional manifestations of the oscillations and to studies of some of their properties, to clarify their nature.

3.2 Lagrangian Radii

Vibrations in Lagrangian radii provide a second manifestation of galactic oscillations. A Lagrangian radius is the radius of a sphere, centered on the mass centroid, that contains a certain fraction of the total mass. Temporal variations in several different Lagrangian radii in experiment 13 are shown in Fig. 3. Counting from the bottom, the tracks show radii that enclose 1/16, 1/8, 1/4, 3/8, 1/2, and 5/8 of the total mass. The top track (outermost of tabulated Lagrangian radii) is similar to that of the kinetic energy (but inverted). It has 15 full cycles during the experiment, and a period of 83 time units. The bottom track (innermost Lagrangian radius) oscillates at higher frequency—34 full cycles during the experiment, for a period of 38 time units.

A curious feature evident in Fig. 3 is the large number of times when all the Lagrangian radii have simultaneous maxima or minima. Simultaneous maxima are marked by vertical solid lines in Fig. 3, while simultaneous minima are marked by vertical dashed lines. These come in an equally curious sequence—two solid lines, then two dashed lines, and so on—and suddenly 3 solid lines. Nearly a pattern, but not quite. Between the solid or dashed lines are times when maxima on the inside (bottom of the plot) turn into minima at the outside (top of the plot) and vice versa. Smooth progressive changes from a maximum to a minimum can be followed as you scan up or down through the tracks.

Simultaneous maxima or minima suggest modes. Waves (or wave packets) travelling inward or outward would give sloping lines connecting maxima or minima. Vertical lines imply infinite phase velocity, which indicates standing waves or modes. If so, two different modes are present—one with period 38 that dominates near the center and another with period around 83 that dominates the outer regions. The two oscillations do not have the low-order commensurability that the large number of simultaneous maxima or minima seem to imply, another property that suggests chaos.

3.3 Modes

Two dominant forms of disturbance are apparent from the results presented so far. Both are spherically symmetrical. Both act like normal modes, so we call them modes. The first mode has no nodes and the second mode has one. The first is a "breathing mode", an homologous expansion or contraction of the entire galaxy. It produces most of the amplitude in the total kinetic energy plots (Fig. 2). We refer to it as the fundamental mode. The second, that with the radial node, is more interesting. We refer to it as the "second radial mode". Hénon (1973) reported this mode some years ago, but he identified it as unstable. While it may have been genuinely unstable in his models, it is also possible that he saw oscillations build up in a stable mode.

The period of the fundamental mode is in good agreement (scatter of $\pm 0.4\%$) with values calculated from the virial equations.

Several other modes showed up as well, but they are much weaker. One has $\ell = 2$ and $m = 0$. The galaxy oscillates between oblate and prolate forms in this mode. It has a period near $46 - 47T$ in experiment 13. An $\ell = 1$ oscillation with a period around 71T is also present in experiment 13. Its period is distinct from that of any of the other disturbances found in that experiment. We looked for other forms of disturbance as well, but with less success. A search for patterns described by spherical harmonics through $\ell = 6$, $m = 6$ showed nothing above the noise, but this search is expensive and it was pursued only over a short segment of one experiment.

It is not trivially obvious that a self-gravitating particle system should show normal modes, but these disturbances act like normal modes. The re-

sults presented here appear to be the first clear-cut experimental evidence for normal mode oscillations in stable particle systems.

Modal amplitudes are much larger than predicted by arguments from statistical mechanics and equipartition, according to which the energy in each mode should be about the average kinetic energy per particle.

The two spherically symmetrical disturbances were confirmed by a separate experiment that was constrained to remain spherically symmetrical. Pseudoparticles are spherical shells in this experiment, which is a kind of N-body calculation that has been used many times (see, e.g., Hénon 1968, 1973). Both disturbances were evident in the results, with the same periods, but the second radial mode was even stronger. This experiment confirms that the disturbances are real—that they are not simply a feature produced by the experimental method.

3.4 Other Properties of Oscillating Galaxies

3.4.1 Modal Structures

There is enough signal in these experiments that we have been able to study Lagrangian displacements. These are changes in Lagrangian radii, the analogue of displacements in fluid dynamics. These displacements show the spatial structures of the fundamental and second modes. Displacements increase linearly with radius in the fundamental mode, while they have a node in the second mode.

3.4.2 Particle Orbits

Orbits of individual particles have been studied as well. Histograms of measured orbital periods show no unusual features near the periods of modes or of their harmonics.

An interesting property of individual particle orbits is that no orbit remained inside the radius of the node in the second radial mode over the duration of the experiment. Every particle ventured outside that radius repeatedly. This result was unexpected so we studied it more carefully, and found that large outer turning point radii are caused by interactions between the second radial mode and individual particle orbits. Some particles would be on regular orbits with small outer turning point radii in the absence of any modal disturbance, but they go chaotic if the amplitude of the second radial mode is large enough (it is in these experiments). And the chaotic form has much larger outer turning point radii.

3.4.3 Non-Self-Consistent Experiment

Self-excited oscillations require some kind of feedback mechanism. The normal-mode oscillations reported here were shown to require a feedback loop to

sustain them by means of a separate, non self-consistent, experiment. Oscillations decay as expected in this experiment. Kinetic energy oscillations are gone by $32T_{cr}$. Distributions of orbital and epicyclic periods for individual stars are similar to those of experiment 13. Again, there are no peaks or dips in the distribution near the periods of known modes.

This experiment demonstrates that the oscillations are maintained by feedback through the self-consistent gravitational potential. An interplay between the particle motions and the self-consistent potential sustains the oscillations.

3.4.4 Do Oscillations Grow Spontaneously?

A question that is important to fundamental understanding of the dynamics of galaxies is whether oscillations are left over from initial conditions or grew spontaneously. The question is not terribly important astronomically since observational tests cannot distinguish one case from the other, but it is important to fundamental understanding.

We have seen examples of both in our experiments. Experiments like 13 and others of its sequence had large amplitude oscillations that clearly were left over from initial conditions, while in some others, experiments for which we had taken some pains to establish very quiet initial conditions, oscillations grew out of the noise. Oscillations that grew out of the noise seldom reached the large amplitudes shown in Fig. 2.

On the other hand, some experiments in the sequence of experiment 13 were started from carefully designed initial conditions that *should* have been steady-state equilibria as defined through the Vlasov equation. The fact that those systems showed large amplitude oscillations (mostly fundamental mode) clearly indicates that the experimental systems do *not* fit the design. They preferred to oscillate about a nearby robust stable equilibrium instead. The problem of finding that nearby robust stable equilibrium analytically is challenging, and is not yet fully appreciated.

This demonstrates that the design of equilibrium starting conditions is a much more delicate process than had been assumed. It also shows a new aspect to experimental stability studies: they system you had in mind might be unstable, but there could be a nearby stable system that is the condition that is actually tested experimentally.

3.5 Observational Consequences

Normal mode oscillations ("ringing") in galaxies can cause observable effects even though their periods are too long ($100-300$ Myr) for direct observation. Ringing brings new features into the interpretation of observations. Galaxies that would otherwise be considered to be in a steady state actually support radial flows. Bulges in spiral galaxies might be a manifestation of normal mode oscillations.

The effects of oscillations are most prominent in spiral galaxies. Most of the mass of the galaxy is presumed to be dark matter, and the luminous material in a spiral galaxy is pretty much confined to a disk (or to a sheet) within the gravitational potential produced by that dark matter. The dark matter "halo" is nearly spherical, and the experimental results reported here apply to it. It is the thing that is oscillating. Luminous material in the disk moves radially in response to driving from those oscillations as it orbits within the basic potential.

3.5.1 Kinematic Warps

The fundamental mode produces flows that are inward during one half-cycle of the oscillation, outward in the next. Combined with the usual rotation within a flat disk, the resulting projected velocity map would be interpreted as indicating a kinematic warp. Flows in the second radial mode would be interpreted as oval distortions. This interpretation of kinematic warps differs substantially from the conventional twisted-rings picture. Radial flows are forbidden in the usual picture of steady state (time independent) potentials, but such flows are present in an oscillating system. Neutral hydrogen in the interstellar medium moves with the fluid dynamical flow indicated by the Lagrangian displacements.

Flow velocities in the fundamental mode are large enough to produce appreciable distortion of the velocity field as seen in 21-cm H I observations and they would combine with the circular motion to be interpreted as a kinematic warp in a map of the velocity field. This flow pattern leads quite naturally to simple symmetries in velocity fields, which would be interpreted as the "hat-brim" warps so common in galaxies. No extra considerations are required to keep them from becoming extremely convoluted, a problem that has always plagued the warp business.

Flows in the second radial mode change sign with radius as well as with the phase of the oscillation. At any instant, they would generate an apparent "warp" that deviates in one sense from purely circular motion in the inner parts of a galaxy, while deviating in the opposite sense farther out. That is the signature of an oval distortion.

We are studying kinematic warps and oval distortions in more detail to see how they relate to normal mode oscillations. We stress that this picture does not help with those warps that are directly observed, such as those seen in edge-on systems and in our own Galaxy. But the picture applies even there: radial inflows and outflows may accompany true geometric warps as well. And the modal picture does not help to explain why directly observed warps are not convoluted.

3.5.2 Bulges

It is tempting to associate bulges in early-type spiral galaxies with the second radial mode. The bulge would result from star formation as a consequence of compression of the normal interstellar medium within the galaxy at times when the phase of that oscillation produces large central densities. It would largely be confined to regions interior to the node. Since the period of the second mode oscillation is on the order of 100 Myr, hot gases could cool, leading to star formation. Most bulges would be observed in stars with ages in that range, but a few might show younger stars.

This fits some of the observed properties of bulges. Their brightness would drop off very rapidly with increasing distance, soon becoming lost in the disk, and appearing consistent with a de Vaucouleurs profile. Bulges would be highly variable from one galaxy to another, both with respect to length scales and to total brightness, even within a given Hubble type. They might be associated with decreased H I emission and with hot gas in the central regions. Their prominence would depend on the amplitude of the second mode oscillation, which can be quite different from galaxy to galaxy. Since bulges would be formed of material gathered from the galactic disk, they would partake of the parent galaxy's rotation or, perhaps, even rotate a bit faster to conserve angular momentum. They would rotate rapidly, rather than slowly like elliptical galaxies. This fits the observations.

3.5.3 Normal Modes

Bulges and kinematic warps might be taken as evidence for normal mode oscillations within galaxies. This merely turns the conjectures of the two previous subsections around. Any galaxy that displays either of these effects is presumed to be ringing. If so, the "halos" supporting most spiral galaxies are oscillating. The dark matter "halos" must be of a dynamical form that supports reasonably large-amplitude oscillations. This provides extra constraints on dynamical models for those halos.

3.6 Concluding Remarks

Oscillations in galaxies are a new research topic. They were a puzzle that we had to track down. We were able to find and to study them because (1) noise was reduced by use of reasonably large numbers of particles and (2) we found a model that supports very large oscillation amplitudes. We concentrated on nonrotating spherical galaxies to simplify both the experimental tools needed and the ease of understanding and interpreting results.

Preliminary reports on this work have already been published (Miller 1991, 1992a). A fuller report has been submitted to *Celestial Mechanics and Dynamical Astronomy.*

Acknowledgements. Computations were made possible by means of a grant of computing services from the NAS division at NASA-Ames, which is gratefully acknowledged. Funds for partial support of this work have been allocated by the NASA-Ames Research Center, Moffett Field, CA, under Cooperative Agreement NCC 2-265 with the University of Chicago. Thanks are also due to the NASA-American Society for Engineering Education for a Summer Faculty Fellowship.

References

Ahmad A., Cohen L., 1973, J. Comp. Phys. 12, 389
Arnold V.I., Avez A., 1968, Ergodic Problems of Classical Mechanics, W. A. Benjamin, Inc., New York
Hénon M., 1968, in: J. Delhaye (ed.), Colloque sur le problème des N corps, p. 241, CNRS, Paris
Hénon M., 1973, A&A 24, 229
Krylov N.S., 1979, Works on the Foundations of Statistical Physics, Princeton University Press, Princeton, N. J.
Lecar M., 1968, Bull. Astron. (Paris) 3, 91
Lorenz E.N., 1963, J. Atmos. Sci. 20, 130
Miller R.H., 1964, ApJ 140, 250
Miller R.H., 1966, ApJ 146, 831
Miller R.H., 1967, J. Comp. Phys. 2, 1
Miller R.H., 1971a, J. Comp. Phys. 8, 449
Miller R.H., 1971b, J. Comp. Phys. 8, 464
Miller R.H., 1974, in: D.G. Bettis (ed.), Proceedings of the Conference on the Numerical Solution of Ordinary Differential Equations, pp 260–275, Springer, Berlin, Vol. 362 of Lecture Notes in Mathematics
Miller R.H., 1991, in: J.R. Buchler, S. Detweiler, and J.R. Ipser (eds.), Nonlinear Problems in Relativity and Cosmology, pp 55–67, Proceedings of Sixth Florida Workshop in Nonlinear Astronomy, Gainesville Fla., 11–13 October 1990, Annals of the New York Academy of Sciences, Vol. 631
Miller R.H., 1992, American Scientist 80, 153
Miller R.H. and Prendergast, K.H., 1968, ApJ 151, 699
Miller R.H., Smith B.F., 1988, in: D.J. Benney, F.H. Shu, and C. Yuan (eds.), Applied Mathematics, Fluid Mechanics, Astrophysics, a Symposium to Honor C. C. Lin, pp 366–372, World Scientific, Singapore
Miller R.H., Smith B.F., 1992, ApJ 393, 508, Fiche 122–B7, includes a videotape
Perdang J.M., 1985, in: J.R. Buchler, J.M. Perdang, and E.A. Spiegel (eds.), Chaos in Astrophysics, pp 11–89, Reidel, Dordrecht
Smith H., 1979, A&A 76, 192
Standish E.M., 1968, Ph.D. thesis, Yale University
von Hoerner S., 1960, Zeitschr. für Astrophys. 50, 184

N-Body Systems: Computer Image and Reality

V.G. GURZADYAN[1] & A.A. KOCHARYAN[2]

[1] Dept. of Theoretical Physics, Yerevan Physics Institute, 375036 Yerevan, Armenia
[2] Dept. of Mathematics, Monash University, Clayton, Victoria, 3168, Australia

Abstract. The problem of the study of N-body gravitating systems is considered from the point of view of the relevance of the computer-created model with the physical one. It is shown that due to the discontinuity of the Lyapunov characteristic exponents their numerical calculations cannot be a proper method of studying the instability properties of the system. This fact occurs both for systems with softened (independent on the way of softening) and, what is more interesting, even for unsoftened (i.e., pure Newtonian) potential.

1. Introduction

For a long time now computer simulations have been one of the important areas of study of the classical problem of N-body gravitating systems. Numerous studies have been performed to simulate different aspects of evolution and structure of galaxies and star clusters. Both the number of those studies and their quantitative characteristics (number of test particles and/or iterations, etc.) are subject to increase, proportionally to the possibilities of the developing computer technique.

During these studies, often without discussion and even without mentioning, a fundamental assumption is being adopted as a quite natural one: the computer created system is the appropriate counterpart of the real system, i.e. the physical system is the limit of the modelled one at smooth change of parameters. Evidently the validity of this assumption, in general, depends not only on the system and the methods of numerical simulations but also on the aims of the study: if in certain cases the computer distortion of the system does not influence at least the main qualitative results, one can have situations when it happens.

Recent new activity of computer studies concerns the instability properties of N-body gravitating systems. The importance of the topic is doubtless, especially now, after the proof of exponential instability of spherical N-body systems [1], and the evidence of radical consequences it can have on the basics

of the dynamics of stellar systems. The validity of the mentioned assumption in this case becomes more crucial, since it is closely related with fundamental concepts such as stability and continuity. Clearly in its general formulation the problem has even a somewhat philosophical context and comes to the relevance of a computer image with reality.

Here we confine ourselves to considering the correspondence of computer studies of instability properties of N-body systems, particularly by means of widely used criteria, to those of physical systems. We approach to this problem from the concepts of theory of dynamical systems, enabling us to arrive to general conclusions, valid for d-dimensional Hamiltonian systems with any potential. Our conclusions do not seem to be quite expected and have even some dramatic shade. In particular, we show that the calculations of Lyapunov exponents for N-body systems with softened and even unsoftened (!) potential can have no relation to the properties of the corresponding real system.

2. Is the function Φ continuous?

Consider a smooth d-dimensional manifold M, with given σ-field of measurable sets $\mathcal{B}(M)$ and complete measure P ($P(M) = 1$). Let f^t be a group of diffeomorphisms on M with continuous $t \in \mathbb{R}$ (or discrete $t \in \mathbb{Z}$) time:

$$f^t : M \to M\,, \qquad f^0 = \mathrm{id}\,, \qquad f^{s+t} = f^s \circ f^t\,, \tag{1}$$

for arbitrary times t, s. Then we will call $(M, \mathcal{B}(M), P, f)$ a dynamical system with continuous (or discrete) time. The dynamical system is measure preserving if

$$P(f^t A) = P(A)\,, \quad \forall A \in \mathcal{B}(M)\,, \quad \forall t \in \mathbb{R}\,. \tag{2}$$

Denote by $\mathcal{D}$ the space of all dynamical systems $(M, \mathcal{B}(M), P, f)$ with appropriate topology. Define the function Φ on $\mathcal{D}$:

$$\Phi : \mathcal{D} \to \mathbb{R}\,. \tag{3}$$

Well known examples of such functions are Kolmogorov-Sinai (KS)-entropy and Lyapunov mean characteristic exponents:

$$\lambda_i = \int\limits_M \lambda_i(x) P(\mathrm{d}x), \quad i = 1, \ldots, q\ , \tag{4}$$

where $\lambda_i(x)$ are Lyapunov characteristic exponents and

$$\lambda_1 > \lambda_2 > \cdots > \lambda_q\,. \tag{5}$$

Now we argue the crucial role of the following question: *Is Φ a continuous function?* At first sight the answer to this question does not seem to be

of much interest, since when one deals with a given dynamical system the behaviour of a certain function for different systems is not important.

It is undoubtedly so for analytical methods, while for computer studies this point can become of extreme importance. Let us illustrate this idea by the following example. Consider the manifold

$$M = S^1 = R^1/Z = \{\theta \mid 0 \leq \theta < 1\} \tag{6}$$

with given Lebesgue measure $P(S^1) = 1$, and class of dynamical systems

$$f_\alpha : S^1 \to S^1 : \theta \mapsto \{\theta + \alpha\}\,, \tag{7}$$

where the brackets denote fractional part. Define Φ in the following way:

$$\Phi(f_\alpha) \equiv \Phi(\alpha) = \begin{cases} 1 & f_\alpha \text{ is ergodic,} \\ 0 & f_\alpha \text{ is not ergodic.} \end{cases} \tag{8}$$

Evidently f_α is ergodic when α is an irrational number, and is non-ergodic when α is rational, i.e.

$$\Phi(\alpha) = \begin{cases} 1 & \alpha \text{ is irrational,} \\ 0 & \alpha \text{ is rational.} \end{cases} \tag{9}$$

Now if one tries to evaluate $\Phi(\sqrt{2})$, one finds out whether the dynamical system $f_{\sqrt{2}}$ is ergodic or not, say, via looking for the periodicity of orbits. Since the computer cannot deal with irrational numbers one is forced to study the following sequence of dynamical systems $\{f_{\alpha_n}\}$ where α_n's are rational numbers for any n and

$$\lim_{n\to\infty} \alpha_n = \sqrt{2}\,. \tag{10}$$

One may expect the following result

$$\Phi(\sqrt{2}) = \lim_{n\to\infty} \Phi(\alpha_n) = 0\,, \tag{11}$$

though real value of $\Phi(\sqrt{2})$ is 1. So

Computer image of $f_{\sqrt{2}} \neq$ Real image of $f_{\sqrt{2}}$.

Therefore one can never figure out by computer methods $\Phi(\alpha)$ when α is irrational. This example clearly shows what in fact happens when one tries to study a dynamical system by computer.

In a typical case one is forced to consider not the diffeomorphisms f^t but

$$f_\varepsilon^t = f^t + \varepsilon(t)\,, \tag{12}$$

because of the inevitable computing errors $\varepsilon(t)$ arising at truncation of numbers. As a result when Φ is not a continuous function all numerical calculations can completely lose their meaning, i.e. computer's $\Phi^c(f)$ does not coincide with real one $\Phi(f)$.

In some cases such a situation can arise as a result of deliberate change of the behaviour of f to avoid "naughty" functions, as it just happens at "softening" of Newtonian potential.

Thus we arrive to the following two questions reflecting both fundamental properties of function Φ:

S. Is $\Phi(\varepsilon)$ close to its computer image $\Phi^c(\varepsilon)$ *(Stability)*?
C. Is $\Phi(0)$ the limit of $\Phi(\varepsilon)$ as ε (bifurcation parameter) tends to zero *(Continuity)*?

As a representation of particular interest of Φ we shall consider Lyapunov characteristic exponents. First, recall the following quite remarkable results:

1a. *According to Mañé's theorem (see* [2]*) when M is a compact surface, C^1 area-preserving non-Anosov diffeomorphisms, all of whose Lyapunov exponents are equal to zero Lebesgue almost everywhere, are everywhere dense;*
1b. *In general, Lyapunov exponents are discontinuous functions of a bifurcation parameter* [3];
1c. *Topological entropy is proved to be discontinuous for* $\dim M \geq 4$ [4].

We see that Lyapunov exponents can be highly discontinuous. The example considered above can be an illustration of this typical property of function Φ. On the other hand one has the properties:

2a. *Topological entropy is continuous at* $\dim M = 2$ [4];
2b. *For some dynamical systems it is proved that Lyapunov exponents are upper-continuous* [5];
2c. *Though typically Lyapunov exponents are highly discontinuous, there exists regular family of perturbations fulfilling conditions discussed in* [2] *making them stable.*

We see that while the properties 1(a-c) make doubtful the usefulness for computations of Lyapunov exponents the 2(a-c) ones leave some hope. What is evident is the necessity of thorough consideration of this problem in any given particular case.

3. N-Body Systems

In view of the results described above let us turn to our problem of the stability in the case of N-body system. First remind that the trajectories of Hamiltonian system

$$H(p,q) = \frac{1}{2} g^{\mu\nu} p_\mu p_\nu + V(q) \tag{13}$$

in the region of configuration space

$$Q = \{q \mid V(q) < E\}\,, \tag{14}$$

can be represented as geodesics of Q with Riemannian metric

$$\mathbf{G} = [E - V(q)]\mathbf{g} \equiv W\mathbf{g}\,, \tag{15}$$

and affine parameter s:

$$\mathrm{d}s = \sqrt{2}W\mathrm{d}t\,. \tag{16}$$

It is also well known that the stability properties of trajectories can be determined by the behaviour of Riemann (Riem), Ricci (Ric) and scalar (R) curvatures. Indeed, Jacobi equation

$$\nabla_{\mathbf{u}}\nabla_{\mathbf{u}}\mathbf{n} + \mathrm{Riem}(\mathbf{n},\mathbf{u})\mathbf{u} = 0\,, \tag{17}$$

where ∇ is the Levi-Civita connection of the metric $\mathbf{G}$, vectors $\mathbf{u}$ and $\mathbf{n}$ denote the velocity on geodesics and their deviation, shows that behaviour of deviation vector $\mathbf{n}$ depends on the Riemannian tensor Riem. Therefore, the problem of continuity of Lyapunov exponents is reduced to the study of the continuity of Riemannian tensor; particularly the latter is not continuous if the scalar curvature R is discontinuous. So the problem comes to the investigation of **instability mean index** λ^2 – a measure of average (in space) instability for the d-dimensional Hamiltonian system:

$$\lambda^2 \equiv -\frac{2RW^2}{d(d-1)} = \frac{2\Delta W}{d} + \left(\frac{1}{2} - \frac{3}{d}\right)\frac{\|\mathrm{d}W\|^2}{W}\,, \tag{18}$$

where the time reparametrization is made and

$$\Delta W = g^{\mu\nu}\frac{\partial^2 W}{\partial q^\mu \partial q^\nu}\,, \qquad \|\mathrm{d}W\|^2 = g^{\mu\nu}\frac{\partial W}{\partial q^\mu}\frac{\partial W}{\partial q^\nu}\,. \tag{19}$$

This quantity is just connected with those Lyapunov exponents which are being obtained during computer calculations.

In the case of N-body gravitating system one has $d = 3N$ and

$$\begin{aligned} V(q) &= -\sum_{a=1}^{N}\sum_{b=1}^{a-1} GM_aM_b\varphi(r_{ab})\,, \\ r_{ab}^2 &= (r_{ab}^1)^2 + (r_{ab}^2)^2 + (r_{ab}^3)^2\,, \qquad r_{ab}^i = r_a^i - r_b^i \end{aligned} \tag{20}$$

where the function φ is not specified yet. Then one has

$$\begin{aligned} &g_{\mu\nu} = M_\mu\delta_{\mu\nu}, \quad \mu = (a,i)\,, \\ &\delta_{\mu\nu} = \delta_{ab}\delta_{ij}, \qquad M_\mu = M_a, \qquad q^\mu \equiv r_a^i \end{aligned} \tag{21}$$

where $a = 1,\ldots,N$ and $i = 1,\ldots,3$. Calculating the instability mean index, taking into account that

$$W = \sum_{a=1}^{N} \frac{|p_a|^2}{2M_a}, \qquad \Delta W = \sum_{a=1}^{N} \sum_{c=1, c\neq a}^{N} GM_c r_{ac}^{-2} \left(r_{ac}^2 \varphi'(r_{ac})\right)',$$
$$\|\mathrm{d}W\|^2 = \sum_{a=1}^{N} G^2 M_a \sum_{i=1}^{3} \left(\sum_{c=1, c\neq a}^{N} M_c \varphi'(r_{ac}) \frac{r_{ac}^i}{r_{ac}} \right)^2 = \sum_{a=1}^{N} \frac{|F_a|^2}{M_a}, \tag{22}$$

one has

$$\lambda^2 = \Lambda_1 + \Lambda_2, \qquad \begin{aligned} \Lambda_1 &= \frac{2}{3N} \sum_{a=1}^{N} \sum_{c=1, c\neq a}^{N} GM_c r_{ac}^{-2} \left(r_{ac}^2 \varphi'(r_{ac})\right)', \\ \Lambda_2 &= \left(\frac{1}{2} - \frac{1}{N}\right) \sum_{a=1}^{N} \frac{|F_a|^2}{M_a} \Bigg/ \sum_{a=1}^{N} \frac{|p_a|^2}{2M_a} \end{aligned} \tag{23}$$

where the following notations are used

$$|F_a|^2 = (F_a^1)^2 + (F_a^2)^2 + (F_a^3)^2,$$
$$F_a^i = \sum_{c=1, c\neq a}^{N} F_{ac}^i = \sum_{c=1, c\neq a}^{N} GM_a M_c \varphi'(r_{ac}) \frac{r_{ac}^i}{r_{ac}}. \tag{24}$$

Now consider a class of potentials $\varphi(\varepsilon)$ containing the two main cases:

1. Newtonian potential ($\varepsilon = 0$): $\varphi(r) = 1/r$. (25)
2. Softened Newtonian potential ($\varepsilon \neq 0$): $\varphi(r, \varepsilon) = 1/\sqrt{r^2 + \varepsilon^2}$. (26)

Let us look for the behaviour of λ when both r and ε are close to zero, i.e. for the continuity of mean index in physically most interesting case. For this purpose one has to obtain the following limits

$$\lim_{\varepsilon \to 0} \lim_{r \to 0} \lambda^2(r, \varepsilon) = -\infty, \qquad \lim_{r \to 0} \lim_{\varepsilon \to 0} \lambda^2(r, \varepsilon) = +\infty. \tag{27}$$

When $\varepsilon = 0$, i.e., in the case of unsoftened potential, the mean index is determined by Λ_2 and the system is exponentially unstable, since

$$\lambda^2 \sim r^{-3} \quad \text{as} \quad r \to 0. \tag{28}$$

This limit corresponds to the close encounter of at least two particles.

The same limit when $\varepsilon \neq 0$, for softened potential reveals completely different behaviour: the mean index is a complex number,

$$\lambda^2 \sim -\varepsilon^{-3} \quad \text{as} \quad \varepsilon \to 0, \tag{29}$$

since it is determined by first member Λ_1; as a result the system is not unstable any more.

We see that the mean instability index and hence Lyapunov exponents in the case of unsoftened potential are discontinuous

$$\lim_{\varepsilon \to 0} \lambda_\varepsilon \neq \lambda_0. \tag{30}$$

Moreover the unsoftened system has quite different behaviour, particularly in accord to point 2b of Sect. 2, it is more stable than the original Newtonian system.

Qualitatively similar situation is in the case of other perturbed potential

$$\varphi_\varepsilon(r) = \frac{1}{r+\varepsilon} \,. \tag{31}$$

Note that dependence of the growth of initial errors on the parameter of softening has been noticed also during computer simulations in [6].

4. Conclusion

Thus the calculations of Lyapunov exponents by means of computer methods for N-body systems can hardly lead to any meaningful results. Already this fact is enough to seriously influence the interpretation of the results of computer studies of instability of softened systems having originated from pioneering paper [7]; see [8] in a way summarising previous studies. Though we did not yet take into account the problems connected with the computer simulation of Lyapunov exponents involving at $t \to \infty$. Other difficulties of those studies, particularly concerning the interpretation of relaxation driving effects, were outlined in [9].

However, the next conclusion of the present study is even more unexpected: the principal impossibility of investigation of instability of not only disturbed, but even $1/r$ potential N-body systems on computers.

These conclusions reflect the cost of inevitable distortion of a system at computer experiments – existence of principal limitations on the obtained information, situation quite familiar from those of quantum mechanical measurements.

References

[1] Gurzadyan V.G., Savvidy G.K., 1984, Doklady AN SSSR 277, 69
[2] Ledrappier F., Young L.-S., 1991, Ergod. Theor. & Dynam. Sys. 11, 469
[3] Ruelle D., 1989, Chaotic Evolution and Strange Attractors, Cambridge U. Press
[4] Katok A., Knieper G., Weiss H., 1991, Commun. Math. Phys. 138, 19
[5] Eden A., Foias C., Temam R., 1991, Journ. Dynam. Diff. Equ. 3, 133
[6] Suto Y., 1991, Publ. Astr. Soc. Jap. 43, L9
[7] Miller R.H., 1964, ApJ 140, 250
[8] Goodman J., Heggie D.C., Hut P., 1992, On the Exponential Instability of N-body systems; submitted to ApJ
[9] Gurzadyan V.G., 1992, in: "Dynamics of Globular Clusters", Proc. Ivan King Workshop, Berkeley

The Approach to Integrability in N-Body Systems with a Central Point Mass

Haywood SMITH [1], Jr., Henry E. KANDRUP [1 2], M. Elaine MAHON [1], & C. SIOPIS [1]

[1] Department of Astronomy, University of Florida, Gainesville, FL 32611, USA
[2] Institute for Fundamental Theory, University of Florida, Gainesville, FL 32611, USA

Abstract. Numerical experiments have been performed with 340-body equal-mass systems having a central point mass. The results indicate that the exponential growth of perturbations in position is partly caused by the "graininess" of the particle distribution, but part is associated with the smooth field of the particles.

1. Introduction

Miller (1964) was the first to show that, if one perturbs the initial positions of particles in a self-gravitating N-body system, the distance in $6N$-dimensional phase space between the perturbed and unperturbed systems grows exponentially with time. Indeed, the perturbation in position for individual particles typically, though not always, grows exponentially. In earlier work (Kandrup & Smith 1991, 1992) the mean e-folding timescale for the growth of the perturbations in position was evaluated for self-gravitating systems and found to be somewhat less than the crossing time $t_{\rm cr}$, defined as $2R_0/v_{\rm rms}$ where R_0 is the initial system radius and $v_{\rm rms}$ is the initial rms velocity of the particles.

The purpose of the experiments described here was to study the approach to integrability in gravitating N-body systems, by introducing a fixed point mass at the center of the system and increasing its mass $M_{\rm c}$ relative to the total mass of the particles $M_{\rm p}$ up to a maximum ratio of 10. In the limit of infinite $M_{\rm c}$, of course, the mean-field potential of the system would be integrable (Kepler potential).

2. Numerical Experiments

The equations of motion for 340 equal-mass particles were numerically integrated using a standard technique described elsewhere (Kandrup & Smith 1991). The potential was the usual softened potential $\phi = Gm_i/(r^2+\epsilon^2)^{\frac{1}{2}}$, most runs having $\epsilon = 10^{-8}$ in units of the initial radius R_0 but a few having

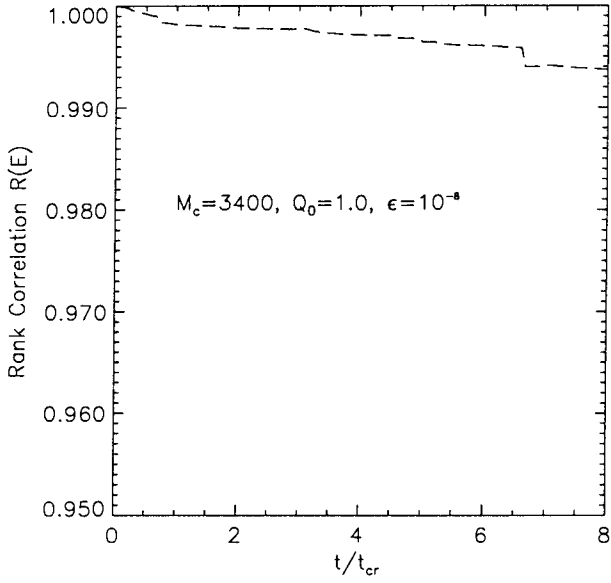


Fig. 1. Rank correlation of energy $R(E)$ vs. time t/t_{cr} averaged over three examples with $M_c = 10M_p$

$\epsilon = 10^{-1.5}$. The central point mass had values M_c of 0, 0.5, 1, 3, 6, and 10 times M_p. The particles were randomly assigned positions with uniform probability density inside a sphere of radius R_0. The particles' velocities were likewise selected randomly with uniform probability density, in most cases with the same maximum speed independent of location but in a few cases with a maximum speed scaled to the local escape speed. All velocities were scaled so as to give a desired virial ratio $Q_0 = 2E_K/|E_P|$, with Q_0 being unity for most experiments but for a few having the value 0.5, giving rise to an initial contraction.

The systems were each integrated for 4 or 8 initial crossing times t_{cr} . For each initial configuration a small ($\delta r_0 = 2 \times 10^{-10}$) random perturbation was added to the position of each particle and the resulting system integrated for the same time as the corresponding unperturbed system. At intervals of $0.05\,t_{cr}$ (for the unperturbed system) the positions and velocities of all particles were stored for later analysis.

3. Results

From the stored output the deviations in position, velocity, and energy (δr_i, δv_i, and δE_i, respectively) were calculated for each particle in the sense (perturbed *minus* unperturbed). From these the total deviations in those same quantities were computed, namely $\Delta r = \frac{1}{2}\ln\sum(\delta r_i^2)$, $\Delta v = \frac{1}{2}\ln\sum(\delta v_i^2)$, and $\Delta E = \frac{1}{2}\ln\sum(\delta E_i^2)$. Additionally, the rank correlation for the particle total energy compared to its initial value, defined as

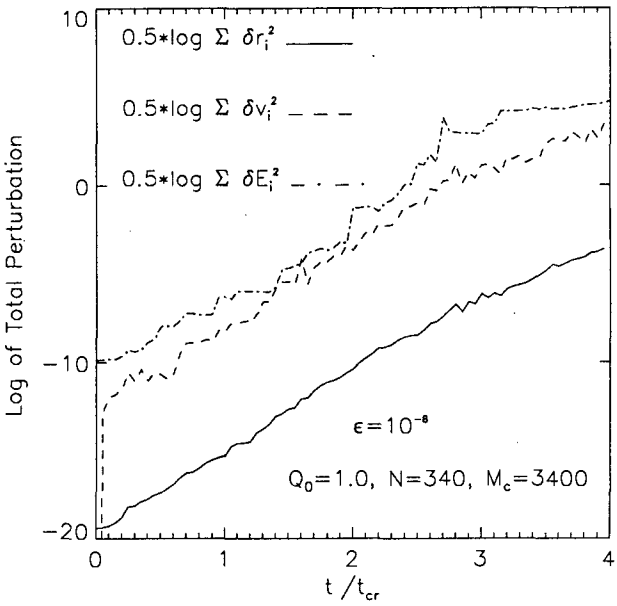


Fig. 2. Total perturbations Δr, Δv, and ΔE as functions of time t for $M_c = 3400$

$$R(E) \equiv 1 - \frac{6}{N^3 - N} \sum_{i=1}^{N} \delta_i^2 \,, \tag{1}$$

was evaluated, where δ_i is the difference in rank order of the energy values for the i^{th} particle. $R(E)$ is a measure of the degree to which the particle Hamiltonian is time-independent.

For the largest value of M_c, namely 3400 ($= 10M_p$), $R(E)$ decreased only very slightly (see Fig. 1), indicating that individual particle energies were very nearly conserved; in the limit they would be conserved exactly. However, for this case (with $Q_0 = 1.0$, $\epsilon = 10^{-8}$) the perturbations in position still grow exponentially, as shown by the total deviation Δr; the same is true for Δv and ΔE as well, as shown in Fig. 2. The growth rate is roughly a factor 3 smaller than for $M_c = 0$. Even the systems having $\epsilon = 10^{-1.5}$ exhibited exponential growth of the total deviations. The growth saturates at later times, when the perturbation has grown to macroscopic size.

The mean e-folding time t_* in units of t_{cr} is shown in Fig. 3 as a function of M_c. While t_* is increasing with M_c, as expected, it is not increasing as rapidly as one might expect. Indeed, from the graph one might be tempted to infer that it is asymptotically approaching a constant value. Eventually it must become very large and the exponential growth disappears, but there is no sign of that in the present data.

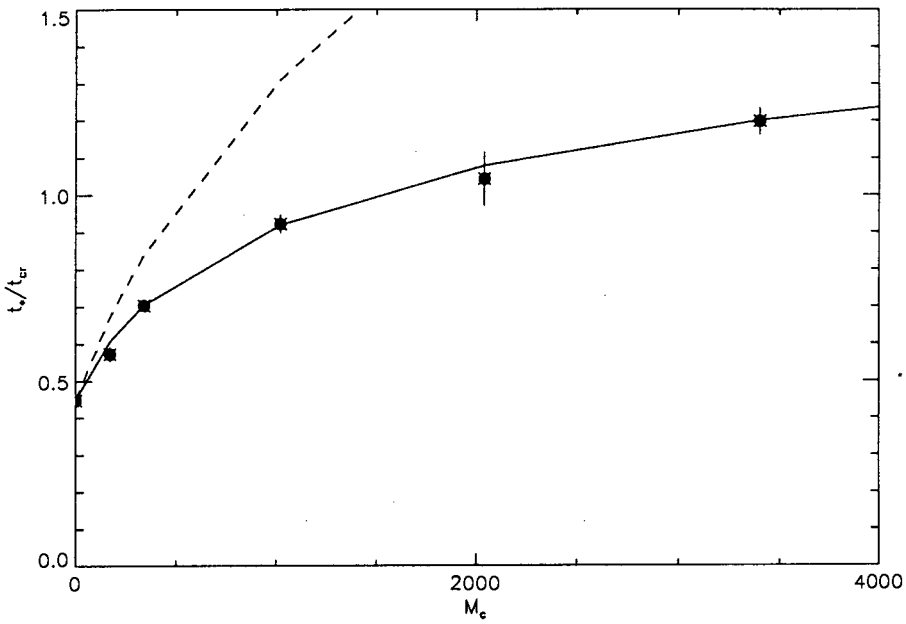


Fig. 3. Mean e-folding time $t_*/t_{\rm cr}$ vs. central mass $M_{\rm c}$. *Dashed curve* is for $t_{\rm gr}$ only (see text); *smooth curve* is for $t_{\rm gr}$ and $t_{\rm sm}$

4. Discussion

The present experiments indicate that exponential instability is persistent in typical particle orbits in gravitating N-body systems up to fairly large values of the ratio $M_{\rm c}/M_{\rm p}$, for which the potential is very nearly Keplerian. It has been argued (cf. e.g., Goodman, Heggie, & Hut 1993) that the sensitivity in these orbits is entirely attributable to the "graininess" of the particle distribution, that is, to the effects of nearby particles. However, if this were true, the e-folding timescale associated with those effects would be essentially proportional to $(Gm/b^3)^{-\frac{1}{2}}$, where m is the particle mass and b is the typical near-neighbor separation. Since in these experiments m and b have not been varied when $M_{\rm c}$ has been changed, $t_{\rm gr}$ should have the same value *in physical units* for all cases. Then, in units of $t_{\rm cr}$, $t_{\rm gr}$ should be inversely proportional to $t_{\rm cr}$, which means that it should be proportional to $(M_{\rm p}+2.5M_{\rm c})^{-\frac{1}{2}}$. (The factor 2.5 appears because of the central concentration of the point mass compared to the distribution of the particles.) The dashed curve in Fig. 3 corresponds to this relation; clearly it does not apply. Similar behavior of $t_*/t_{\rm cr}$ with increasing $M_{\rm c}$ was found with the experiments having $Q_0 = 0.5$.

Consider the smoothed potential associated with the particles: it is almost certainly asymmetric on fairly large scales because of the finite N. It therefore seems likely that this potential is non-integrable. If so, the perturbations

arising from the "graininess" will grow exponentially because of the smoothed potential. Pfenniger (1986) demonstrated a similar effect in a barred galaxy model into which was introduced a fixed perturbing mass; the perturbation in position grew exponentially (on average) at a rate independent of the mass of the perturber up to a fairly large (10^{-2}) mass ratio. If there are *two* exponentiating processes we might then expect that

$$\delta r = \delta r_0 \cdot \mathrm{e}^{t/t_*} = \delta r_0 \cdot \mathrm{e}^{\alpha t} \cdot \mathrm{e}^{\beta t} \,, \tag{2}$$

where $\alpha = 1/t_{\rm gr}$ and $\beta = 1/t_{\rm sm}$, with $t_{\rm sm}$ being the e-folding time appropriate to the effect of the smooth field. We would assume that $t_{\rm sm}$ is a constant in units of $t_{\rm cr}$, because the latter is the basic timescale appropriate to the smoothed field; let us call that constant C. Then, after multiplying by $t_{\rm cr}$, we have

$$D\,\frac{t_{\rm cr}}{t_*} = \frac{t_{\rm cr}}{t_{\rm gr}} + C^{-1} \,. \tag{3}$$

The solid curve in Fig. 3 is a graph of such a relation for $C = 2.02$; as may be seen it is a remarkably good fit to the data. The ratio of the two timescales, $t_{\rm sm}/t_{\rm gr}$, is approximately 4.5 when $M_{\rm c} = 0$. A comparison of systems with a highly softened potential (in which the near-neighbor effects are presumably suppressed) with unsoftened systems, neither having a central point mass, reveals that the mean e-folding time increases by roughly a factor of 3 (Kandrup & Smith 1991); this fact implies that the ratio of the two timescales is at least 2 (which must be a lower limit because the softening does not completely eliminate near-neighbor effects). Thus we conjecture that the stochasticity in N-body systems is largely, but not entirely, due to neighboring particles, this component diminishing smoothly in relative importance as $M_{\rm c}/M_{\rm p}$ is increased; the remainder seems to be caused by asymmetries in the smoothed potential of the particles.

References

Goodman J., Heggie D.C., Hut P., 1993, preprint
Kandrup H.E., Smith H., jr., 1991, ApJ 374, 255
Kandrup H.E., Smith H., jr., 1992, ApJ 386, 635
Miller R.H., 1964, ApJ 140, 250
Pfenniger D., 1986, A&A 165, 74

On the Non-Trivial Concept of Relaxation in N-body Systems

Piero CIPRIANI [1 2] & Giuseppe PUCACCO [2 3]

[1] Scuola di Dottorato di Ricerca in Fisica,
[2] International Center for Relativistic Astrophysics, University of Rome "La Sapienza", I-00185 Roma, Italy
[3] II University of Rome "Tor Vergata", I-00133 Roma, Italy

1. Introduction

There is no doubt that to speak about relaxation towards equilibrium in N-body self-gravitating systems is not an easy task. The difficulties are related, first of all, to the elusivity, in this context, of the very meaning of *equilibrium*; consequently, one must content oneself with the study of the processes through which the system forgets progressively some memory of the initial conditions. So, one has not to describe the relaxation to equilibrium, but, rather, to trace several histories, the plots of which have different players performing on different time-scales.

We surely know of two of these processes: the first one is the establishing of the dynamical equilibrium; the second (but we could equally well say the last) corresponds to "something similar" to the approach to the canonical equilibrium of statistical mechanics. It is a common opinion[4],[5] that in order to reach the dynamical equilibrium (starting from an arbitrary initial bounded state) the system needs a time of the order of the crossing time ($\tau_C \sim R/\sigma$, where R is a characteristic length scale and σ is the RMS velocity), whereas to reach equipartition of energy it needs a time of the order of $N\tau_C/\ln N$ ($\sim \tau_B$, the binary relaxation time-scale). One more point of consensus is that if a system tends to forget its initial conditions on a certain time-scale (say τ_S), it must also display a chaotic behaviour, namely its dynamics must be extremely sensitive to small changes in initial condition. A fundamental question is therefore the following: is τ_S related to τ_C or to τ_B or to some other (eventually intermediate) time-scale? The main lines of thinking on this topic are two: **a)** $\tau_S \sim \tau_C$ (a result obtained by Kandrup[15],[16],[17] and by Goodman, Heggie & Hut[12], with different approaches and also with a different interpretation as far as the sources of instability); **b)** $\tau_S \sim N^{1/3}\tau_C$ (Gurzadyan & Savvidy[13]).

The aims of the present contribution are to point out that the relationships between the *instability time-scale,* as measured from the exponential divergence of trajectories, both from numerical integrations and from "averaged" analytical estimates, the *mixing time*, if it can be estimated in some way, and the *"relaxation" time* (when it is possible to speak about of one relaxation time) are in general highly non-trivial.

So, a very cautious treatment of these concepts is needed in order to avoid some contradictions and misunderstandings, which, in our opinion are mainly due to some unjustified extrapolations and approximations made in (most of) the treatments presented in the last years.

Clearly, the most self-consistent approach to the problem outlined above should be properly carried out in the framework of general Hamiltonian many-body dynamics, and this has already been done[21] for a class of many degrees of freedom Hamiltonians of interest mainly in solid state physics, and working in a slightly different manifold. The work along this line is going[7],[8]; nevertheless our concern here is to point out that most of the conclusions drawn until now, concerning in particular the N-body self-gravitating problem, should be analysed carefully, and that the contradictions between analytical estimates based on different approximations, and the results found by numerical simulations are due to a set of excessive simplifications of the geometro-dynamical quantities governing the evolution of perturbations, on one side; and also to an inadequate representation of the dynamics, on the other side.

In particular, if we can observe that the semi-analytical estimate of the "collective relaxation time-scale", made by Gurzadyan & Savvidy[13], is based on some unrealistic schematizations, nevertheless, the general setting of the problem and the references to the rigorous results of ergodic theory are stated in a nearly correct way.

On the other hand, it is also certainly true that the time-scale of instability (i.e. of "exponential growth") of perturbations in a N-body self-gravitating systems is of the order of the crossing time, but, in principle, this *is not necessarily related* with any kind of relaxation, and in particular, with the *mixing* property of dynamical systems.

2. Crossing and Collective Relaxation Time-Scales

To test the sensitivity of a system to changes in initial conditions the first simple approach is to compare the evolutions followed by the same system starting from a set of different, but very close, initial conditions: let us consider, for simplicity, two points in the $6N$-dimensional phase space of the system and, choosing in some appropriate way a "distance" in the phase space, let us trace it in time as the two trajectories evolve from the initial points. If the system dynamics is sensitive to initial conditions, this distance

grows very fast (indeed at an exponential rate). The problem with this simple approach is that the two compared trajectories explore different regions of phase space with a very different structure[21]: stated in another way the problem is that we have two distinct systems with no further knowledge of the fate of each other. A technique devised to overcome this difficulty is that of following the evolution of the perturbation to a reference system, namely to study the tangent dynamics: a version of this technique (due to Miller[20]) is exploited by Goodman, Heggie & Hut[12]. But, even with this approach, the same problem is encountered, since the solution of the variational equations attains quickly a non-linear regime. Together with these troubles there is another question, common to both the above approaches, concerning the reliability of extrapolations of numerical results: with $N \sim 10^3$ the mean and fluctuating fields are of comparable amplitude; so the techniques used, which are not able to disentangle the effects due to these distinct contributions, make predictions on the rate of instability that are dominated by the more unstable bunch of degrees of freedom, which are not necessarily representative of the average behaviour of the whole system. In addition, there exists a much more awkward problem raised by the interpretations of numerical N-body simulations performed until now: if these were intended to estimate something like the Maximum Lyapunov exponent (MLE) of the dynamical system describing the model of a stellar system, it is very embarrassing to remember that, even for systems with few degrees of freedom, the time of integration needed to obtain a reliable estimate of the MLE can be unimaginably long[21],[9], whereas N-body simulations stop usually before ten crossing times. Again, even when accurately determined, the reciprocal of the MLE has in general no direct and obvious relation with the relaxation time-scale; but on this point, at least, now it seems to exist a general consensus. On the other hand, the work of Gurzadyan & Savvidy[13] addressed from the very beginning to the role played by fluctuations in determining the instability; but, at the same time, underestimated the effect of the mean bulk field[15],[16] and the peculiar structure of a realistic self-gravitating system, with its enormous density contrast and the strong anisotropy of the velocity distribution. On this basis, the averages performed in the configurations and velocity spaces are to be taken only as a first rough estimate of the instability growth rate.

In the light of the above considerations it appears clear that to devise a reliable technique for the description of the instability we need a tool suitable to describe the detailed evolution of the perturbation as a manifestation of a global instability. This tool is the Jacobi–Levi-Civita equation of geodesic deviation on the configuration manifold.

3. The Jacobi–Levi-Cìvita Equation of Geodesic Deviation

The framework in which we settle for the study of the long term behaviour of N-body systems is that of exploiting the equivalence of the trajectories of a conservative Hamiltonian system and the geodesics of the Riemannian space with the corresponding Jacobi metric. Using the minimum action principle in the form given by Maupertuis and Jacobi, one can show[2],[14],[22] that the geodesic equation associated to the metric

$$\begin{aligned} ds^2 = g_{ab}\, dq^a\, dq^b &= [E - V(\mathbf{q})]\, \eta_{ab}\, dq^a\, dq^b \\ &(a, b = 1, \ldots 3N; \quad \eta_{ab}\eta^{bc} \equiv \delta_a^c)\,, \end{aligned} \tag{1}$$

$$\frac{Du^a}{ds} \equiv \frac{du^a}{ds} + \Gamma^a_{bc} u^b u^c = 0\,; \tag{2}$$

and whose explicit form for the velocity vector $u^a = dq^a/ds$ is

$$\frac{du^a}{ds} + \frac{1}{2(E-V)} \left(g^{ab} - 2u^a u^b\right) V_{,b} = 0\,, \tag{3}$$

where with D/ds we indicated the covariant derivative of a vector field along the flow, coincides with the canonical equations with Hamiltonian function

$$H = \tfrac{1}{2}\eta^{ab} p_a p_b + V(\mathbf{q}) \equiv E\,; \tag{4}$$

where E is the total energy of the system, $V(\mathbf{q})$ is the potential energy, and

$$\begin{aligned} &q^a \equiv \sqrt{m^a}\, x^a\,, \qquad m^{3i-2} = m^{3i-1} = m^{3i}\,, \\ &\mathbf{r}_i \equiv \left(x^{3i-2}, x^{3i-1}, x^{3i}\right)\,, \end{aligned} \qquad i = 1, \ldots, N\,. \tag{5}$$

In this approach the stability properties of the dynamics are therefore described by the Jacobi–Levi-Cìvita (JLC) equation of geodesic deviation

$$\frac{D^2 \delta q^a}{ds^2} + R^a{}_{bcd} u^b \delta q^c u^d = 0\,, \tag{6}$$

where δq^a is the perturbation vector in the positions and $R^a{}_{bcd}$ is the Riemann curvature tensor associated to g_{ab}. An equivalent form of the JLC is, more simply,

$$\frac{D^2 \delta q^a}{ds^2} + \mathcal{R}^a{}_c \delta q^c = 0\,, \tag{7}$$

where the stability matrix

$$\mathcal{R}^a{}_c = R^a{}_{bcd} u^b u^d \tag{8}$$

has been introduced. The solutions of the JLC equation allow to predict the evolution of perturbations[21],[7],[8] and to identify the geometric properties of

the manifold related to the instability behaviour of the dynamics. One of the interesting and immediate results of this approach, when applied to the case of the dynamical ("violent") relaxation of self-gravitating N-body systems, consists in the possibility of identifying, in a completely analytical way, a transition in the global stability properties of the dynamics, at the onset of the virial equilibrium, put forward by a damping of the fluctuations of the trace of the stability matrix (about a positive definite average value) which, through the solutions of JLC equation, indicates a transition towards a less unstable behaviour, at least on a global scale.

In almost all preceding studies[13],[15],[16],[12],[21] the analysis was based on the study of the JLC equation for δq^a.

4. The Eigenvalues of the Jacobi–Levi-Cìvita Equation

In the light of the above discussion it is clear that, in order to follow the subsequent evolution of the system we have to come back to the complete JLC Eq. (7). In particular, are of great importance the eigenvalues $\{\rho_{(k)}\}$ of the stability matrix and defined by the equation:

$$\mathcal{R}^a{}_c v^c_{(k)} = \rho_{(k)} v^a_{(k)}, \quad k = 0, \ldots, 3N-1\,, \tag{9}$$

with $\rho_{(o)} = 0$, and $v^a_{(o)} = \mathcal{P}u^a$.

Those eigenvalues, linked, through the JLC equation, to the evolution of the covariant derivatives of the components of perturbations, and, as it is easy to see, to the effective growth of perturbations[7],[8], give informations equivalent to that provided by other typical stochasticity indicators studied in dynamical systems theory, such as the Lyapunov exponents, but in a more manageable form. The reader can refer to the studies in [21],[7],[8] for a full account of this approach, but here we will try to describe the main issues.

As told before, the most direct way to gain some more accurate information about the global behaviour of perturbations, is to consider the equation for the components. As shown elsewhere ([7],[8]), this equation allows to extract many of the features of the dynamical system, and in particular, for a N-body self-gravitating system, the accurate analysis of the information contained in it, shows how peculiar this dynamical system is, and how much significant is to speak about *multiple (instability) time-scales in stellar dynamics*: the analysis of the eigenvalues of the stability matrix, shows the existence, for a system of N self-gravitating point masses, of a full hierarchy of time-scales, whereas, for the usual Hamiltonian systems with many degres of freedom there exists a generic homogeneity of time-scales.

The explicit expression for $\mathcal{R}^a_c$:

$$\mathcal{R}^a_c = \left[\frac{1}{4W^2}\left(\frac{\mathrm{d}W}{\mathrm{d}s}\right)^2 - \frac{1}{2W}\frac{\mathrm{d}^2W}{\mathrm{d}s^2}\right]\delta^a_c + \frac{(\overrightarrow{\nabla W})^2}{4W^3}u^a u_c$$
$$+ \frac{1}{2W}\left(W_{,bc}\,u^a u^b + g^{ab}W_{,bd}\,u^d u_c - g^{ab}W_{,bc}\right) \tag{10}$$
$$- \frac{3}{4W^2}\left[\left(\frac{\mathrm{d}W}{\mathrm{d}s}\right)(u^a W_{,c} + g^{ab}W_{,b}\,u_c) - g^{ab}W_{,b}\,W_{,c}\right],$$

where $W \equiv W(\mathbf{q}) \stackrel{\text{def}}{=} E - V(\mathbf{q})$, allows us to affirm that, also for the evolution of the components of the perturbation, the behaviour will change when the virialization is completed.

It is also possible to show[6],[7],[8] that the spectrum of eigenvalues of the stability matrix $\mathcal{R}$ describes in a coherent way the evolutionary properties of the system in the full phase space, giving a very effective tool for the determination of the stochastic character of the motion. Moreover, the study of the statistical behaviour of Hamiltonian systems with many ($N \gg 3$) degrees of freedom, passing by the study of the geometrical properties of the manifold on which motion takes place, leads to the assessment of possible transitions in the chaotic regime in dependence to the binding energy of the systems, a property which is generic to most of the well studied ones.

5. Conclusions and Results

Two observations to conclude: as shown by Pettini[21], the geometro-dynamical properties of the configuration manifold of a Hamiltonian system with many degrees of freedom are very rapidly variable functions of the point (and then of time). This implies that the numerical simulations cannot have *any immunity* at all with respect to the linearly-induced exponential growth[6],[7],[8] of perturbations due to the different values of curvature (for example) experimented by two different realizations of the same system. More: it is possible (with high probability) that the exponential growth of *errors* in N-body systems is related not (or not only) to the negativity of the curvature, but (also) to the parametric instability induced by the rapidly fluctuating geometrical properties: if this is the case, there is no rigorous result which connect the instability with "mixing".

In order to speak of multiple relaxation time-scales (as it is possible to speak about multiple instability time-scales) it is necessary to show that some mechanism is effective in driving the system toward an alternative and intermediate kind of equilibrium, and that this happens on a time-scale different by some order of magnitude from any other characteristic time, see e.g. [19],[10],[7],[8],[6].

The main issues of our studies are described in [7],[8].

References

[1] Anosov D.V., 1967, Proc. Steklov Inst. Math. 90
[2] Arnold V.I., 1989, Mathematical Methods of Classical Mechanics (IInd Edition), Springer-Verlag
[3] Benettin G., Galgani L., Strelcyn J.M., 1976, Phys.Rev. A 14-n.6, 2338
[4] Binney J., Tremaine S., 1987, Galactic Dynamics, Princeton Univ. Press, Princeton
[5] Chandrasekhar S., 1960, Principles of Stellar Dynamics, Dover
[6] Cipriani P., Ph.D.Thesis, Univ. of Rome
[7] Cipriani P., Pettini M., Pucacco G., (in preparation)
[8] Cipriani P., Pucacco G., 1993, Il Nuovo Cimento B, to appear
[9] Contopoulos G., Barbanis B., 1989, Astron. & Astroph. 222, 329
[10] Galgani L., 1988, in Nonlinear Evolution and Chaotic Phenomena, G. Gallavotti, P.F.Zweifel (eds.), NATO-ASI Series:B 176, Plenum, p. 147-159
[11] Gaspard P. 1992, Preprint Université Libre de Bruxelles
[12] Goodman J., Heggie D.C., Hut P., 1992, Preprint Princeton
[13] Gurzadyan V.G., Savvidy G.K.,1986, Astron. & Astroph. 160, 203
[14] Jacobi C.G.J., 1884, Vorlesungen über Dynamik, Reiner, Berlin, (IInd edition Chelsea, New York, 1969)
[15] Kandrup H.E., 1989, Phys.Lett. A 140, 97
[16] Kandrup H.E., 1990, ApJ 364, 420
[17] Kandrup H.E., Smith H., 1991, ApJ 374, 255
[18] Krylov N.S., 1979, Works on Foundations on Statistical Physics, Princeton Univ. Press
[19] Ma K.S., 1988, Statistical Mechanics, World Scientific
[20] Miller R.H., 1971, J. Comp. Phys. 8, 449
[21] Pettini M., 1993, Phys.Rev. E 47, 828, and these Proceedings
[22] Synge J.L., 1960, On the geometry of dynamics, Handbuch der Physik, III-1, Springer

Gravothermal Oscillations

Rainer SPURZEM

Institut für Theoretische Physik und Sternwarte der Universität Kiel, Germany

Single mass spherical star clusters are considered as idealized model systems for globular clusters or galactic nuclei. Two-body relaxation by distant encounters drives them secularly through a sequence of hydrostatic equilibria, because the relaxation time-scale usually is much longer than the dynamical time-scale. Energy transport within the system was believed to lead for point-mass systems in finite time to infinite central density ("gravothermal catastrophe", Lynden-Bell & Wood 1968; Hachisu et al. 1978). The effect can be understood as a consequence of the negative specific heat in the core of self-gravitating systems. However, at large central densities, the evolution is not any longer dominated by small angle distant two-body encounters. Close three-body encounters produce high-energy escapers as well as an energy input into various zones of the cluster due to their reaction products (Spitzer & Mathieu 1980; Giersz & Spurzem 1993, henceforth GS). Consistently there are post-collapse models of globular clusters with a central energy source, which was seen as a first approximation for the energy input due to hard encounters (Hénon 1965; Inagaki & Lynden-Bell 1983).

Sugimoto & Bettwieser (1983) and Bettwieser & Sugimoto (1984) detected in isotropic conducting gas sphere models with a distributed energy source in the core large amplitude non-linear oscillations, which they called "gravothermal oscillations". Goodman (1987) showed that such oscillations are caused by an instability or overstability of a self-similar post-collapse solution. Stable post-collapse models only exist if the energy generation is larger than a certain threshold, which corresponds in the simplified single mass model to a particle number of $N < 7000$.

Gravothermal oscillations were later also found in direct numerical solutions of the orbit-averaged Fokker-Planck equation (Cohn et al. 1989). Numerical parameter studies with such models showed that the unstable cases are extremely sensitive to numerical errors in their long-term post-collapse evolution and that the dynamical evolution can be described by a low-dimensional attractor, which appears in several cases to be chaotic (Breeden et al. 1992, henceforth BPC).

It is not clear to what extent gravothermal oscillations can be observed in real astrophysical star clusters. In most models core collapse is reversed at a very small core particle number – thus any energy generation should occur in the real N-body system as a stochastic process instead of the smooth, thermodynamic energy generation assumed in the statistical models. Takahashi & Inagaki (1991) showed in their Fokker-Planck models, however, that stochastic energy generation does not change the global picture of gravothermal oscillations. A collisional $N = 10^4$ calculation has been processed until "core bounce" (Spurzem & Aarseth 1993; see also some Figs. in GS); in pre-collapse there is fair agreement with the statistical model based on the Fokker-Planck approximation. To judge about post-collapse, however, the integration time is not yet sufficient.

For the models presented in this paper the Boltzmann equation with a Fokker-Planck collisional term is considered as the fundamental kinetic equation. It is appropriate for systems whose evolution is dominated by distant binary encounters; since it is an evolution equation for the one-particle distribution function any correlations, as e.g. formation and evolution of binaries can not be included at this point. A set of moment equations is derived from the kinetic equation up to second order and closed by a specially tailored heat flux equation in third order, which describes the heat transport by distant binary encounters. In the isotropic case such approach is well known (Lynden-Bell & Eggleton 1980; Heggie 1984) – the moment equations are just conservation equations for mass, momentum and energy. The closure equation is similar to that used e.g. in models of stellar evolution (but with another time-scale in the conductivity); therefore such models are commonly denoted as gaseous models of star clusters. Here it is taken into account that in spherical symmetry the radial and tangential "thermal" energy (i.e. the radial and tangential velocity dispersions) are not in equipartition in real star clusters since the time-scale for collisional isotropization is very long compared to the dynamical time-scale. Therefore in addition to the isotropic gaseous model there are a correction in hydrostatic equilibrium, two separate energy equations and two closure equations for transport of the radial and tangential energy (in radial direction) and in second order a collisional term describing the decay of anisotropy.

Let the dependent variables be the mass M_r contained in a sphere of radius r, the local mass density ρ, radial and tangential pressure p_r, p_t, bulk mass transport velocity u, and transport velocities v_r, v_t of the radial and tangential energy, respectively. As auxiliary quantities the radial and tangential 1-D velocity dispersions $\sigma_r^2 = p_r/\rho$, $\sigma_t^2 = p_t/\rho$, the average velocity dispersion $\sigma^2 = (\sigma_r^2 + 2\sigma_t^2)/3$, the anisotropy $A = 2 - 2\sigma_t^2/\sigma_r^2$ and the relaxation time

$$T = \frac{9}{16\sqrt{\pi}} \frac{\sigma^3}{G^2 m \rho \log(\gamma N)} \tag{1}$$

in the definition of Larson (1970) are used, where N is the total particle number of the star cluster, m the individual stellar mass and γ a numerical

constant whose value will be discussed below. The equations are

$$\frac{\partial M_r}{\partial r} = 4\pi r^2 \rho \,, \tag{2}$$

$$\frac{\partial \rho}{\partial t} + \frac{1}{r^2}\frac{\partial}{\partial r}\left(\rho u r^2\right) = 0 \,, \tag{3}$$

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial r} + \frac{GM_r}{r^2} + \frac{1}{\rho}\frac{\partial p_r}{\partial r} + 2\,\frac{p_r - p_t}{\rho r} = 0 \,, \tag{4}$$

$$\frac{\partial p_r}{\partial t} + \frac{1}{r^2}\frac{\partial}{\partial r}\left(p_r u r^2\right) + 2\,p_r\frac{\partial u}{\partial r} + \frac{3}{r^2}\frac{\partial}{\partial r}\left(p_r(v_r - u)r^2\right) - 4\,\frac{p_t(v_t - u)}{r} = -\frac{2}{3}\frac{p_r - p_t}{\lambda_A T_A} + \left(\frac{\delta p_r}{\delta t}\right)_{\mathrm{bin3}} \,, \tag{5}$$

$$\frac{\partial p_t}{\partial t} + \frac{1}{r^2}\frac{\partial}{\partial r}\left(p_t u r^2\right) + 2\,\frac{p_t u}{r} + \frac{1}{r^2}\frac{\partial}{\partial r}\left(p_t(v_t - u)r^2\right) + 2\,\frac{p_t(v_t - u)}{r} = \frac{1}{3}\frac{p_r - p_t}{\lambda_A T_A} + \left(\frac{\delta p_t}{\delta t}\right)_{\mathrm{bin3}} \,, \tag{6}$$

$$v_r - u + \frac{\lambda}{4\pi G\rho T}\frac{\partial \sigma^2}{\partial r} = 0 \,, \tag{7}$$

$$v_r = v_t \,. \tag{8}$$

The net transport velocities for radial and tangential energy $(v_r - u)$ and $(v_t - u)$ can be derived from the energy fluxes F_r and F_t (which are identified with the third order moments of the velocity distribution) by dividing out a convenient multiple of the relevant pressure ($2p_t$ for $(v_t - u)$, $3p_r$ for $(v_r - u)$). The reader interested in more details about this and the connection of the variables to moments of the stellar velocity distribution is referred to Louis & Spurzem (1991, henceforth LS).

The numerical constants λ_A, λ and γ occurring in Eqs. (5) to (7) are related to the time-scales of collisional anisotropy decay and heat transport, and to the Coulomb logarithm, respectively. λ is related to the standard C constant in isotropic gaseous models (see e.g. Heggie & Stephenson 1988) by

$$\lambda = \frac{27\sqrt{\pi}}{10} C \,. \tag{9}$$

T_A is the anisotropy decay time-scale for an anisotropic local velocity distribution function; for a generalization of Larson's (1970) distribution function (series of Legendre polynomials) including anisotropy it is $T_A = 10\,T/9$ (see LS; in contrast to the isotropic result $5\,T/6$, which has been frequently used in other work). Since the real distribution function is not completely determined within the framework of the gaseous model a free numerical constant λ_A is introduced in Eqs. (5) and (6).

Finally, the additional terms due to the average heating by formation and hardening of three-body binaries (see e.g. Cohn 1985) is

$$\left(\frac{\delta p_r}{\delta t}\right)_{\text{bin3}} = \frac{2}{3} C_b \frac{\rho^3}{m^2\sigma^2}\left(\frac{Gm}{\sigma}\right)^5, \tag{10}$$

$$\left(\frac{\delta p_t}{\delta t}\right)_{\text{bin3}} = \left(\frac{\delta p_r}{\delta t}\right)_{\text{bin3}}. \tag{11}$$

This is an isotropic energy input. It is shown in GS and Giersz & Heggie (1993ab) that for particle numbers between $N = 10^3$ and $N = 10^4$ the best agreement between direct N-body calculations, direct solutions of the orbit-averaged Fokker-Planck equation and this anisotropic gaseous models is achieved for one set of parameters, namely $\lambda = 0.4977$ (i.e. $C = 0.104$), $\gamma = 0.11$, $\lambda_A = 0.1$, and $C_b = 90$. Note that the last value is the standard value derived from theoretical reasoning and three-body experiments, and it is independently confirmed by GS, and that the value of γ found empirically is somewhat smaller than e.g. Spitzer's (1987) standard value ($\gamma = 0.4$).

Strictly the initial ansatz using the one-particle distribution function without any correlations except distant binary encounters is invalid in the presence of binaries. However, there appear to be only very few binaries in the core of a cluster, which only at short time intervals of maximum central density influence the dynamical evolution; moreover one can judge empirically from GS that such description of the binary effects as phenomenological heating term included afterwards in the second order moment equations is a fair approximation in a statistical sense to what happens in the real N-body system.

For the numerical solution of the model equations and comparisons with direct N-body results standard N-body units were used, where $G = 1$, the total mass of the system $M = 1$, and the total energy of Plummer's model, which was used as initial model, is $E = -1/4$; in these units the scaling radius of Plummer's model is $a = 3\pi/16$. The equations were discretized on a Eulerian mesh with 200 logarithmically equidistant grid points; they were distributed between a minimal and maximal radius of $2.06 \cdot 10^{-6}$ and 144. in the above units. A partially implicit Newton-Raphson-Henyey method was used to solve the difference equations. The numerical results were tested by comparing with the known self-similar solutions of LS.

As a function of particle number as control parameter there are two critical numbers; first N_1 marking the transition from a steady regular post-collapse expansion at low N to overstable post-collapse solutions exhibiting regular oscillations, the other N_2 marking the transition to unstable large amplitude gravothermal oscillations, presumably of chaotic nature, as discussed for the isotropic case by Goodman (1987) and BPC. BPC also show that in their numerical models there are some periodic windows even for $N > N_2$. However, for the anisotropic case the corresponding work has not yet been done; we are currently performing as a first step a numerical survey to approximately determine the critical numbers and the behaviour of the post-collapse models (Spurzem & Louis 1993). But we are well aware that it is very difficult to numerically prove the existence of chaos.

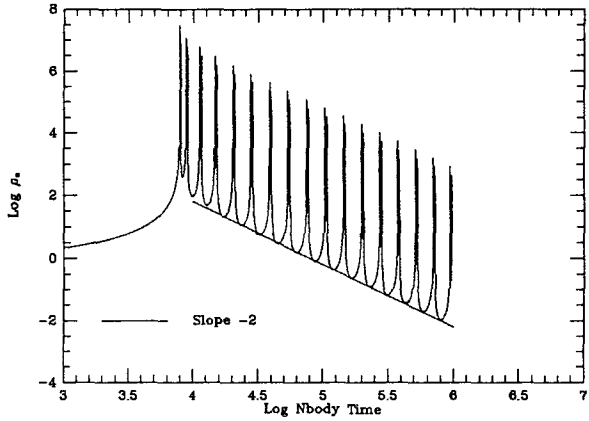


Fig. 1. Central density of $N = 5 \cdot 10^4$ star cluster vs. time in N-body units ($G = M = 1$, $E = -1/4$)

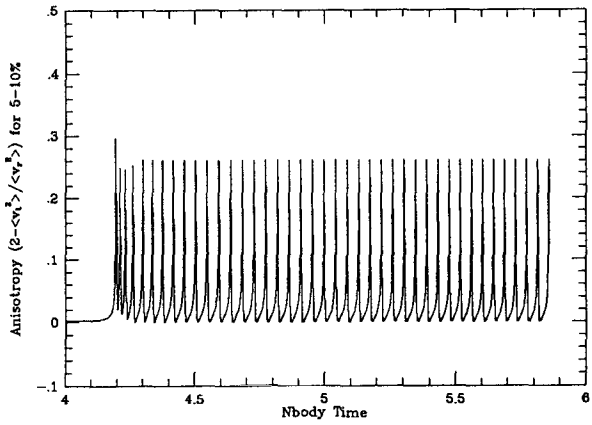


Fig. 2. Anisotropy as a function of N-body time averaged over the Lagrangian mass shell containing $5 - 10\%$ of the total mass

As in the isotropic Fokker-Planck models a low-dimensional attractor describes well the deterministic evolution of the system. Such result is plausible in the light of results from Allen & Heggie (1992), who have demonstrated that many features of gravothermal oscillations can be simulated by dynamical equations of a simple three-zone model. The variables sufficient to describe the dynamical evolution of the cluster core are "detrended" values of central density ρ_c', central velocity dispersion σ_c', and a dimensionless core collapse rate $\xi = t_{rx} d \ln \rho_c / dt$, where t_{rx} is the local two-body relaxation time-scale. Fig. 1 shows the evolution of ρ_c for a $N = 5 \cdot 10^4$ model. "Detrending" a quantity means here to divide it by a global power law t^α (whose value α is taken from Goodman's (1987) self-similar post-collapse solutions). It is for example $\alpha = -2$ for the central density such that $\rho_c' = t^2 \rho_c$ has a constant time average. It is remarkable that here with anisotropy the same power-laws can be applied to get constant time averages as in the isotropic case of Goodman (1987). Thus for, e.g., $N = 10^5$ a very nice periodic attractor

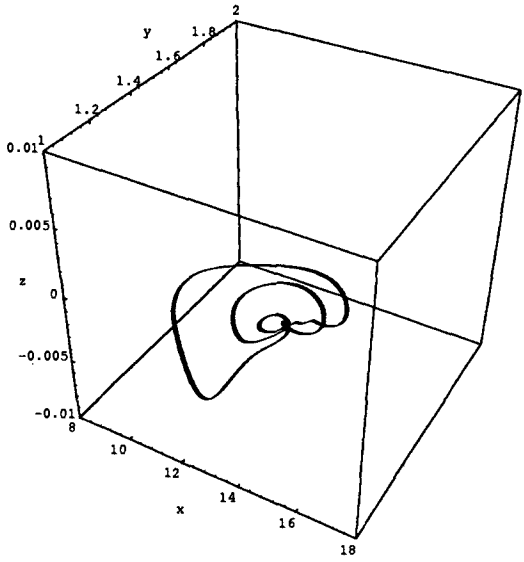


Fig. 3. Projected three-dimensional attractor for a $N = 10^5$ system, $x = \log \rho_c'$, $y = \log \sigma_c'$, $z = \xi$

was found, whose projection is shown in Fig. 3. Fig. 2 depicts an example of the evolution of anisotropy in one of the central Lagrangian mass shells for another model.

The results obtained so far show compared to BPC for the anisotropic gaseous model a larger fraction of regular periodic attractors. From the present results it cannot be excluded that there also may exist chaotic solutions for some particle numbers. The dynamical evolution of the presumably chaotic solutions is similar to the Rössler attractor (Jackson 1987, see also BPC). Future work will present these numerical results in more detail (Spurzem & Louis 1993); furthermore the critical numbers N_1 and N_2 should be derived for the anisotropic case, which requires first to derive new self-similar anisotropic post-collapse models with a distributed energy source, because previous models including anisotropy (LS) are singular ones.

Acknowledgements: I thank P.D. Louis (Munich) for close cooperation; some of the results presented here will be part of a common paper with him to be published. Very fruitful cooperation and communication with M. Giersz and D.C. Heggie (Edinburgh) is gratefully acknowledged. Computations have been done at CRAY YMP HLRZ Jülich and KSR1-32 of GWD Göttingen.

References

Allen F.S., Heggie D.C., 1992, MNRAS, 257, 245
Bettwieser E., Sugimoto D., 1984, MNRAS, 208, 493
Breeden J.L., Packard N.H., Cohn H.N., 1992b, Chaos in Astrophysical Systems, preprint
Cohn H., 1985, in Goodman J., Hut P., eds, Proc. IAU Symp. 113, Dynamics of Star Clusters, Reidel, Dordrecht, p. 161
Cohn H., Hut P., Wise M., 1989, ApJ, 342, 814
Giersz M., Spurzem R., 1993, subm. to MNRAS, (GS)
Giersz M., Heggie D.C. 1993a, subm. to MNRAS
Giersz M., Heggie D.C. 1993b, to be subm. to MNRAS
Goodman J., 1987, ApJ, 313, 576
Hachisu I., Nakada Y., Nomoto K., Sugimoto D., 1978, Prog. Theor. Phys., 60, 393
Heggie D.C., 1984, MNRAS, 206, 179
Heggie D.C., Stephenson D., 1988, MNRAS, 230, 223
Hénon M., 1965, An d'Ap, 28, 62
Inagaki S., Lynden-Bell D., 1983, MNRAS, 205, 913
Jackson E.A., 1990, Perspectives of nonlinear dynamics 2, Cambridge Univ. Press, Cambridge
Larson R.B., 1970, MNRAS, 147, 323
Louis P.D., Spurzem, R., 1991, MNRAS, 244, 478
Lynden-Bell D., Eggleton P.P., 1980, MNRAS, 191, 483
Lynden-Bell D., Wood R., 1968, MNRAS, 138, 495
Spitzer L., 1987, Dynamical Evolution of Globular Clusters, Princeton Univ. Press, New Jersey, USA
Spitzer L., Mathieu R.D., 1980, ApJ, 241, 618
Spurzem R., Aarseth S.J., 1993, in prep.
Spurzem R., Louis P.D., 1993, to be subm. to MNRAS
Sugimoto D., Bettwieser E., 1983, MNRAS, 204, 19p
Takahashi K., Inagaki S., 1991, PASJ, 43, 589

Recent Results on the Stability of Anisotropic Stellar Systems

Jérôme PEREZ [1 2] & Jean-Jacques ALY [1]

[1] Service d'Astrophysique - CE Saclay - 91191 Gif sur Yvette Cédex - France,
[2] ETCA/CREA - 16 bis av. Prieur de la Côte d'Or - 94114 Arcueil Cédex - France

Abstract. We present a new approach to deal with the problem of stability of collisionless stellar systems. This technique, based on the symplectic structure of the Vlasov-Poisson system, allows us to derive a new stability criterion for general systems. It is very useful in the anisotropic spherical case.

1. Introduction

As remarked by Zwicky in the thirties, owing to their physical aspect galaxies and comparable stellar systems seem to be relaxed, but their collisional relaxation time is in general larger than the age of the universe. So, there exists an important collisionless epoch in the life of stellar systems which governs a part of their dynamical evolution. Taking into account the statistical behaviour of their components, stellar systems can be treated by a Kinetic Theory. However, as stressed by several authors (see Binney & Tremaine, 1987 – hereafter BT – for a review) classical stability techniques stops when the problem becomes inhomogeneous and anisotropic. In this sense, following the pioneer works of Bartholomew (1971) reactualized by Kandrup (1991a), we present here a new symplectic approach of dynamical evolution of collisionless stellar systems. This approach permits to obtain a quite simple and very general stability criterion, which, in the special case of spherical isopotential curves allows us to generalise an important isotropic result called sometimes Antonov-Lebovitz Theorem.

2. A Symplectic Stability Criterion

As shown by Morrison (1982), the standard Vlasov equation can be expressed by using functionals. Indeed, if F is any functional involving the phase space, the motion evolution equation can be written

$$\dot{F}[f] = \int f \left[\frac{\delta F}{\delta f}, \frac{\delta H}{\delta f}\right] \mathrm{d}\Gamma := \{F, H\}[f] \,, \tag{1}$$

where $[,]$ denotes the standard Poisson bracket in the canonical conjugated variables p and q, $\{,\}$ has all the properties of a Lie bracket, and $\delta/\delta f$ means a functional derivative operation.

Owing to the fact that any physical perturbation of any initial state f_0 can be described by a generator g which is a Hamiltonian-like function representing the effectuated canonical transformation (e.g., Bartholomew 1971), we can find a general Taylor expansion for any functional F during this perturbation

$$\begin{aligned} \hat{F}[f_0] &= F[f_0] - \{G, F\} + \tfrac{1}{2}\{G, \{G, F\}\} - \tfrac{1}{3!}\Big\{G, \{G, \{G, F\}\}\Big\} + \dots \\ &= \left(e^{\{\cdot, G\}} F\right)[f_0] = F\left[e^{[g, \cdot]} f_0\right] \,, \end{aligned} \tag{2}$$

where the functional generator G is such that $\delta G/\delta f = g$.

It is important to note that the development in Eq. (2) is true for any functional, so we can apply it to all the functionals intervening in our problem such as the energy, the entropy or more complicated ones.

Applying Eq. (2) to the special case of the total energy of the system

$$H[f] = \int \mathrm{d}\Gamma \, \frac{p^2}{2m} \, f(q, p, t) \; - \; \frac{Gm^2}{2} \int \mathrm{d}\Gamma \int \mathrm{d}\Gamma' \, \frac{f(q, p, t)\, f(q', p', t)}{|q - q'|} \,, \tag{3}$$

and choosing for f_0 a steady state, the first order energy variation is clearly vanishing, and the second order one can be written

$$H^{(2)}[f_0] = -\frac{1}{2} \int [g, E][g, f_0] \, \mathrm{d}\Gamma \; - \; \frac{Gm^2}{2} \int \mathrm{d}\Gamma \int \mathrm{d}\Gamma' \, \frac{[g, f_0].[g', f_0']}{|q - q'|} \,, \tag{4}$$

where E is the single-particle energy, functional derivative of $H[f]$. The stability of the system against some perturbations generated by some g, can now be investigated in the anisotropic inhomogeneous case by the study of the sign of $H^{(2)}[f_0]$. However, this development gives us only a sufficient condition for stability. Indeed, we cannot make, as Laval, Mercier, & Pellat (1965), a connection between $H^{(2)}[f_0]$ and some definite inner product to have a general energy principle. As quoted by Larson (1991) the inner-product related to our general problem is indefinite, and in this case relations like Schwartz's inequality fails and does not allow us to connect our energy variation to some dynamical variable to diagnostic linear instabilities. However, as illustrated in Kandrup (1991b), negativity of $H^{(2)}[f_0]$ can imply secular instability, and a way is open in this weak version of instability.

3. Stability Results

As quoted in BT, the stability problem of a collisionless stellar system and of a gaz volume in gravitational interaction are closely related. In fact, in many cases the hydrodynamic problem (when identifiable) is simpler than its stellar analogue. Indeed, a hydrodynamic problem is a three dimensional one, while stellar systems have six degrees of freedom. Hence a technique to simplify the study of stellar systems consists in averaging over velocities, when it is possible to deal with the hydrodynamic counterpart of the problem. One of the most important isotropic result, sometimes called Antonov-Lebovitz Theorem (ALT) (see BT), is built on this technique and assures the stability of $f_0(E)$'s systems against non-radial perturbations.

In a recent paper (Aly & Perez 1992), we present a new demonstration of this important result. This is the combination of this new method and the symplectic approach of the stability criterion which allows us to obtain our result on the stability of spherical stellar systems.

Before to give the proof we have to install the context and to define the class of perturbations which allows us to use the technique.

A stationary spherical system has a distribution function which depends only on the energy E and on the squared norm L^2 of the angular momentum L, $f_0 = f_0(E, L^2)$. We assume here that

$$f_E := \frac{\partial f_0}{\partial E} \leq 0\,, \qquad \text{and} \qquad f_{L^2} := \frac{\partial f_0}{\partial L^2} \leq 0\,. \tag{5}$$

We want to consider the stability of such an equilibrium with respect to the class of preserving perturbations generated by all the functions g satisfying $[g, L^2] = 0$. It is easy to see what is precisely the class of preserving perturbations by writing the definition of a g-generated perturbation

$$f_1 = [g, f_0(E, L^2)] = f_E[g, E] + f_{L^2}[g, L^2]\,. \tag{6}$$

This equation is written in the general anisotropic spherical case. In order to clarify the situation, it is clear that in the isotropic case where f_0 depends only of E, all the perturbations are preserving because g always commutes with L^2. Moreover, the stability of isotropic systems in the conditions of Eq. (5) is well known. On the other hand, in the case of purely radial orbit systems, with distribution function of the form $f_0(E, L^2) = \phi(E)\cdot\delta(L^2)$ (e.g., Fridman & Polyachenko 1984), where ϕ is any smooth function and δ denotes the Dirac distribution, one can show that there exists no non-vanishing g such that $[g, L^2] = 0$ and $f_1 \neq 0$ simultaneously. In this sense we say that purely radial orbit systems cannot receive preserving perturbations.

In the intermediate case, preserving perturbations are generated by all the functions of the form $g(L^2, L_x, L_y, L_z)$, for example all the radial perturbations are also preserving. Non-radial and preserving perturbations are clearly

mathematically defined and numerically seen as not forming a vanishing set, but we haven't seen their physical interpretation.

In the general spherical case one can split the perturbation f_1 into two parts, with one part being invariant under rotations,

$$\begin{aligned} f_1 &= \overline{f_1} + \delta f_1 \,, \\ \overline{f_1}(q,p) &= \int f_1(R(q),R(p))\,\mathrm{d}R, \quad \text{and} \quad \int \delta f_1(R(q),R(p))\,\mathrm{d}R = 0 \,, \end{aligned} \tag{7}$$

where the averaging is made other all possibles rotations R. Hence, one can show that the second order variation of the energy splits into $H^{(2)}[f_1] = H^{(2)}[\overline{f_1}] + H^{(2)}[\delta f_1]$. The first part being positive by Kandrup & Sygnet (1985) Theorem, we study the second non-radial part and show (using the revisited proof of ALT and the well known Wirtinger inequality)

$$H^{(2)}[\delta f_1] \geq \frac{1}{2}\int \frac{(\delta\rho_1)^2}{|f_E|}\,\mathrm{d}\mathbf{r} - \frac{Gm^2}{2}\int \mathrm{d}\mathbf{r}\int \mathrm{d}\mathbf{r}'\,\frac{\delta\rho_1\delta\rho_1'}{|q-q'|} \geq 0 \,, \tag{8}$$

where the non-radial part of the perturbed density $\delta\rho_1$ is directly obtained from δf_1 by a velocity averaging. This result assures the stability of any anisotropic spherical stellar system having the properties in Eq. (5) against preserving perturbations. Therefore, it should be stressed that the preserving character is not robust dynamically. To have a greater signification this stability result must be completed by a numerical experiment concerning the existing relation between the nature of perturbation affecting a wide class of anisotropic spherical systems and their preserving character. These kinds of simulations are under investigation and preliminary results tend to confirm that in most cases, stable systems are preserving-perturbed, and unstable systems are non-preserving perturbed (Perez et al. 1993; Perez & Alimi 1993).

References

Aly J.J., Perez J., 1992, MNRAS 259, 95
Bartholomew P., 1971, MNRAS 151, 333
Binney J., Tremaine S., 1987, Galactic Dynamics, Princeton Univ. Press.
Fridman A.M., Polyachenko V.L., 1984, Physics of Gravitating Systems, New York, Springer
Kandrup H.E., 1990, Astrophys. J. 351, 104
Kandrup H.E., 1991a, Astrophys. J. 370, 312
Kandrup H.E., 1991b, Astrophys. J. 380, 511
Kandrup H., Sygnet J.F., 1985, Astrophys. J. 298, 27
Laval G., Mercier C., Pellat R., 1965, Nuclear Fusion 5, 156
Larson J., 1991, Physics Review Let. 66, 1466
Morisson P.J., 1980, Physics Let. A 80, 383
Perez J., Alimi J.M., Aly J.J., Scholl H., 1993, Proceedings of the Gravitational Meeting of Aussois, Springer, in press
Perez J., Aly J.J., 1993, MNRAS, in preparation
Perez J., Alimi J.M., Aly J.J., Scholl H., 1993, MNRAS, in preparation

5. The Few-Body Problem

The Stability of the Solar System

Jacques LASKAR

Astronomie et Systèmes Dynamiques, Bureau des Longitudes, 3 rue Mazarine, 75006, Paris, France

The problem of the stability of the solar system has fascinated astronomers and mathematicians since antiquity, when it was observed that among the seemingly fixed stars, there were also "wandering stars"—the planets. Efforts were first focused on finding a regularity in the motion of these wanderers, so their movement among the fixed stars could be predicted. For Hipparcus and Ptolemy, the ideal model was a combination of uniform circular motions, the epicycles, which were continually adjusted over the centuries to conform to the observed course of the planets. Astronomy had become predictive, even if its models were in continual need of adjustment.

From 1609 to 1618, Kepler fixed the planets' trajectories: having assimilated the lessons of Copernicus, he placed the Sun at the center of the universe and, based on the observations of Tycho Brahe, showed that the planets described ellipses around the Sun. At the end of a revolution, each planet found itself back where it started and so retraced the same ellipse. Though seductive in its simplicity, this vision of a perfectly stable solar system in which all orbits were periodic would not remain unchallenged for long.

In 1687 Newton announced the law of universal gravitation. By restricting this law to the interactions of planets with the Sun alone, one obtains Kepler's phenomenology. But Newton's law applies to all interactions: Jupiter is attracted by the Sun, as is Saturn, but Jupiter and Saturn also attract each other. There is no reason to assume that the planets' orbits are fixed invariant ellipses, and Kepler's beautiful regularity is destroyed.

In Newton's view, the perturbations among the planets were strong enough to destroy the stability of the solar system, and divine intervention was required from time to time to restore planets' orbits to their place. Moreover, Newton's law did not yet enjoy its present status, and astronomers wondered if it was truly enough to account for the observed movements of bodies in the solar system.

The problem of solar system stability was a real one, since after Kepler, Halley was able to show, by analyzing the Chaldean observations transmitted by Ptolemy, that Saturn was moving away from the Sun while Jupiter was

moving closer. By crudely extrapolating these observations, one finds that six million years ago Jupiter and Saturn were at the same distance from the Sun. In the 18th century, Laplace took up one of these observations, which he dated March 1st, 228 BC: *At 4:23 am, mean Paris time, Saturn was observed "two fingers" under Gamma in Virgo.* Starting from contemporary observations, Laplace hoped to calculate backward in time using Newton's equations to arrive to this 2000 year-old observation.

The variations of planetary orbits were such that, in order to predict the planets' positions in the sky, de LaLande was required to introduce artificial "secular" terms in his ephemeris tables. Could these terms be accounted for by Newton's law?

The problem remained open until the end of the 18th century, when Lagrange and Laplace correctly formulated the equations of motion. Lagrange started from the fact that the motion of a planet remains close, over a short duration, to a Keplerian ellipse, and so had the notion to use this ellipse as the basis for a coordinate system. Lagrange then wrote the differential equations that govern the variations in this elliptic motion under the effect of perturbations from other planets, thus inaugurating the methods of classical celestial mechanics. Laplace and Lagrange, whose work converged on this point, calculated secular variations, in other words long-term variations in the planets' semi-major axes under the effects of perturbations by the other planets. Their calculations showed that, up to first order in the masses of the planets, these variations vanish (Poisson and Poincaré later showed that this result remains true through second order in the masses of the planets, but it is not valid at the third order).

This result seemed to contradict Ptolemy's observations from antiquity, but by examining the periodic perturbations between Jupiter and Saturn, Laplace discovered a quasi-resonant term $(2\lambda_{\text{Jupiter}} - 5\lambda_{\text{Saturn}})$ in their longitudes. This term has an amplitude of 46′50″ in Saturn's longitude, and a period of about 900 years. This explains why observations taken in 228 BC and then in 1590 and 1650 could give the impression of a secular term.

Laplace then calculated many other periodic terms, and established a theory of motion for Jupiter and Saturn in very good agreement with 18th century observations. Above all, using the same theory, he was able to account for Ptolemy's observations to within one minute of arc, without additional terms in his calculations. He thus showed that Newton's law was in itself sufficient to explain the movement of the planets throughout known history, and this exploit no doubt partly accounted for Laplace's determinism.

Laplace showed that the planets' semi-major axes undergo only small oscillations, and do not have secular terms. At the same time, the eccentricity and inclination of planets' trajectories are also very important for solar system stability. If a planet's eccentricity changes appreciably, its orbit might cut through another planet's orbit, increasing the chances of a close encounter which could eject it from the solar system.

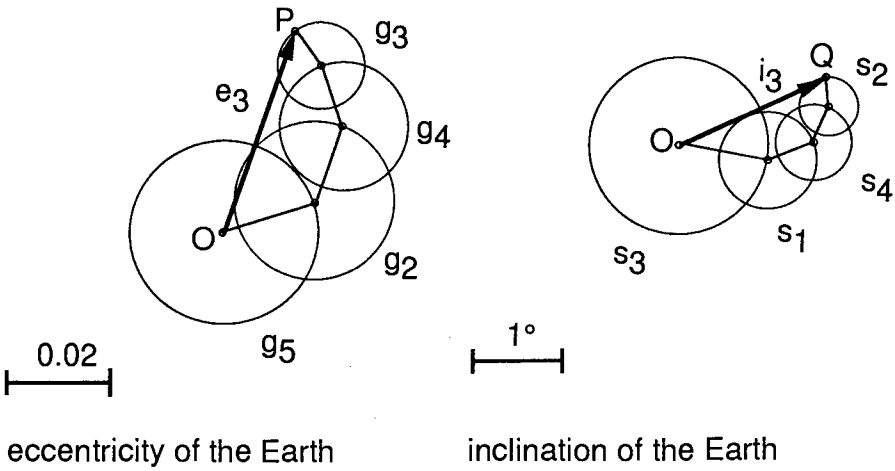


Fig. 1. The solutions of Laplace for the motion of the planets are combinations of circular and uniform motions with frequencies the precession frequencies g_i and s_i of the solar system, which correspond to periods from about 50000 years to several million years. The eccentricity e_3 of the Earth is given by OP, while the inclination of the Earth with respect to the invariant plane of the solar system (i_3) is OQ (Laskar 1992b)

Laplace revisited his calculations, taking into account only terms of first order in the perturbation series, and showed that the system of equations describing the mean motions of eccentricity and inclination may be reduced to a system of linear differential equations with constant coefficients. He also showed, using the conservation of the angular momentum, that the solutions of this system are quasi-periodic (linear combinations of periodic terms), and that the variations in eccentricity reduce to a superposition of uniform circular motions (Fig. 1). The inclinations and eccentricities of the orbits are therefore subject to only small variations about their mean values. But it must be stressed that Laplace's solutions are very different from Kepler's, because the orbits are no longer fixed. They are subject to a double precessional motion with periods ranging from 50,000 to several million years: precession of the perihelion, which is the slow rotation of the orbit in its plane, and precession of the nodes, which is the rotation of the plane of the orbit in space.

Later, Le Verrier, famed for his discovery in 1846 of the planet Neptune through calculations based on observations of irregularities in the movement of Uranus, took up Laplace's calculations and considered the effects of higher order terms in the series. He showed that these terms produced significant corrections and that Laplace's and Lagrange's calculations "could not be used for an indefinite length of time". He then challenged future mathematicians to find exact solutions, without approximations. The difficulty posed by "small divisors" showed that the convergence of the series depended on

initial conditions, and the proof of the stability of the solar system remained an open problem.

Between 1892 and 1899 Poincaré formulated a negative response to Le Verrier's question. In so doing he rethought the methods of celestial mechanics along the lines of Jacobi's and Hamilton's work. In his memoir "On the three body problem and the equations of dynamics," Poincaré showed that it is not possible to integrate the equations of motion of three bodies subject to mutual interaction, and not possible to find an analytic solution representing the movement of the planets valid over an infinite time interval, since the series used by astronomers to calculate the movement of the planets were not convergent.

In the 1950's and 60's, the mathematicians Kolmogorov, Arnold, and Moser took up Poincaré's work and showed that, for certain values of the initial conditions, it was nonetheless possible to obtain convergent series. If the masses, eccentricities, and inclinations of the planets are small enough, then many initial conditions lead to quasi-periodic planetary trajectories. But the actual masses of the planets are much too large for this result (known as the KAM theorem) to apply directly to the solar system and thereby prove its stability.

All the efforts since the first demonstration of Laplace, were devoted to the search of proofs of stability for the solar system. But as Poincaré already forecast, all these proofs were only approximations, and none of them prevented the solar system for being unstable on very long time scales. The proof of Arnold (1963), is a rigorous proof, but as it is stated above, it does not apply directly to our solar system to demonstrate its stability. Anyway, all these elements contributed to the general feeling that the solar system was stable, and this question was not much investigated, although the problem remained unsolved.

But in the past five years, the understanding of the solar system stability question has advanced considerably, partly due to the improvement of computer technology.

One part of the efforts towards an understanding of the long time behavior of the solar system consists of direct numerical integration of the equations of motion (Newton's equations, sometimes with additional relativistic corrections or perturbations due to the Moon). Initial studies were limited to the motion of the outer planets, from Jupiter to Pluto. In fact, the more rapid the orbital movement of a planet, the more difficult it is to numerically integrate its motion. To integrate the orbit of Jupiter, a step-size of 40 days will suffice, while a step-size of 0.5 days is required to integrate the motion of the whole solar system. The project LONGSTOP (Carpino et al. 1987; Nobili et al. 1989) must be counted among the recent studies; this used a CRAY to integrate the system of outer planets over a model time of 100 million years. At about the same time, calculations of the same system were carried out at MIT over even longer periods, corresponding to times of 210 and 875 mil-

lion years. These calculations were carried out on the "Orrery," a vectorized computer specially designed for the task (Applegate et al. 1986; Sussman & Wisdom 1988). This latter integration showed that the motion of Pluto is chaotic, with a Liapunov exponent of 1/20 million years. But since the mass of Pluto is very small, (1/130 000 000 the mass of the Sun), this does not induce macroscopic instabilities in the rest of the solar system, which remained stable in these numerical studies.

My approach was different, and more in the spirit of the analytical works of Laplace and Le Verrier. Indeed, since these pioneer works, the Bureau des Longitudes, has traditionally been the place for development of analytical planetary theories (Brumberg & Chapront 1973; Bretagnon 1974; Duriez 1979). All these studies are based on classical perturbation series. Implicitly, they assume that the motion of the celestial bodies they study are regular and quasi periodic. The methods used are essentially the same which were used by Le Verrier, with the additional help of the computers. Indeed, such methods can provide very satisfactory approximations of the solutions of the planets over a few thousand years, but they will not be able to give answers to the question of the stability of the solar system over time span comparable to its age. This difficulty is the main reason which motivated the long time direct numerical integrations of the gravitational equations. This new approach, which was made possible by the use of the computers, can give very precise solutions of the trajectories, but they are limited by the short step-size necessary for the integration of the whole solar system. It should be stressed, that until 1991, the only numerical integration of a realistic model of the full solar system was the numerically integrated ephemeris DE102 of JPL (Newhall et al. 1983) which spanned only 44 centuries.

At first, I tried to extend as far as possible the classical analytical planetary theories, but I realized quite rapidly that it was hopeless when considering the whole solar system, because of severe convergence problems encountered in the system of the inner planets (Laskar 1984). I thus decided to proceed in two very distinct steps: a first one, purely analytical is the averaging of the equations of motion over the rapid angles, that is the motion of the planets along their orbits. This process was conducted in a very extensive way, without neglecting any term, up to second order with respect to the masses, and through degree 5 in eccentricity and inclination. The system of equations thus obtained comprises some 150,000 terms, but it can be considered as a simplified system, as its main frequencies are now the precessing frequencies of the orbits of the planets, and no longer comprise their orbital periods. The full system can thus be numerically integrated with a very large step-size of about 500 years. Contributions due to the Moon and to general relativity are added without difficulty.

This second step, i.e. the numerical integration, is very efficient because of the symmetric shape of the secular system, and was conducted over 200 million years in just a few hours on a super computer. The main results of this

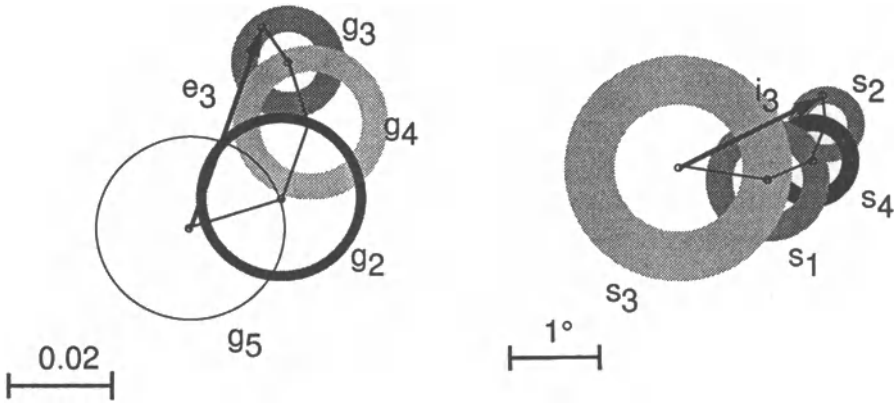


Fig. 2. Solutions for the eccentricity and inclination of the Earth. The effect of the chaotic dynamics in the solar system can be pictured by enlarging each circle of Fig. 1, which corresponds to a fundamental frequency of the solar system, into a band corresponding to the chaotic zone. The position of the junction points inside this zone will be unpredictable after several million years (Laskar 1992b)

integration was to reveal that the whole solar system, and more particularly the inner solar system (Mercury, Venus, Earth, and Mars), is chaotic, with a Liapunov exponent of 1/5 million years (Laskar 1989). An error of 15 m in the Earth's initial position gives rise to an error of about 150 m after 10 million years; but this same error grows to 150 million km after 100 million years. It is thus possible to construct ephemerides over a 10 million year period, but it becomes practically impossible to predict the motion of the planets beyond 100 million years.

This chaotic behavior essentially originates in the presence of two secular resonances among the planets: $\theta = 2(g_4 - g_3) - (s_4 - s_3)$, which is related to Mars and the Earth, and $\sigma = (g_1 - g_5) - (s_1 - s_2)$, related to Mercury, Venus, and Jupiter (the g_i are the secular frequencies related to the perihelions of the planets, while the s_i are the secular frequencies of the nodes) (Laskar 1990). The two corresponding arguments change several times from libration to circulation over 200 million years, which is also a characteristic of chaotic behavior.

When these results were published, the only possible comparison was the comparison with the 44 centuries ephemeris DE102, which already allowed to be confident on the quality of the results (Laskar 1986, 1990), but at the time, I thought it would need about 10 years before similar results could be obtained with direct numerical integration of the gravitational equations. In fact, due to the very rapid advances in computer technology, and in particular to the development of workstations, only two years later, Quinn et al. (1991) published a numerical integration of the full solar system, including

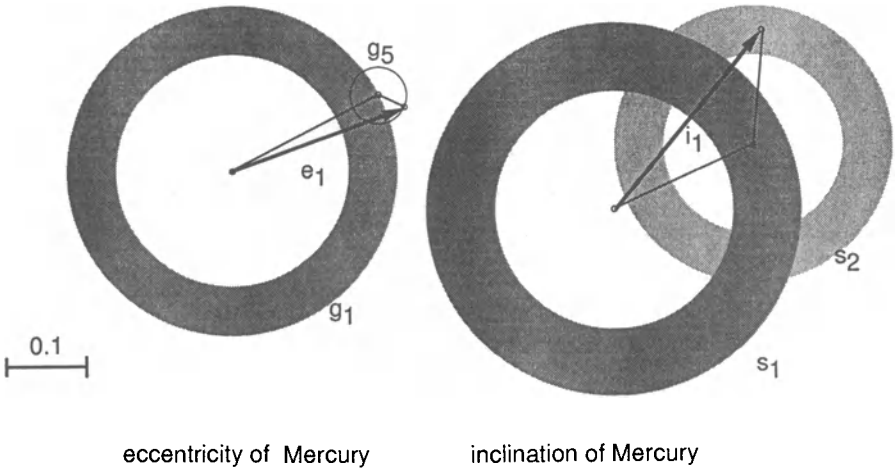


Fig. 3. Eccentricity and inclination of Mercury (Laskar 1992b)

the effects of general relativity and the Moon, which spanned 3 million years in the past (completed later on by an integration from −3 Myr to +3 Myr). Comparison with the secular solution (Laskar 1990) shows very good quantitative agreement, and confirms the existence of secular resonances in the inner solar system (Laskar et al. 1992a). Later on, using mapping techniques, Sussman & Wisdom (1992) published an integration of the solar system over 100 million years which confirms the existence of the secular resonances as well as the value of the Liapunov exponent for the solar system.

The solar system, and more particularly the inner solar system, is strongly chaotic, but on a rather long time scale. This is because the fundamental periods of the motions involved are the precessional periods of the orbits (on the order of 100,000 years), and not the orbital periods (on the order of a year). Over periods on the order of 100 million years, the variations of planetary eccentricities and inclinations are dominated by the quasi-periodic components already present in the solutions of Laplace and Le Verrier. The effects of chaos over 400 million years may nevertheless be estimated to about 0.01 for the Earth's eccentricity and one degree for its inclination (Fig. 2). These are of course only lower estimates, since we do not yet have the means to bound these variations from above over several hundred million years, which would require a more global knowledge of the phase space of motion.

The most perturbed planet is Mercury, the effects of its chaotic dynamics being clearly visible over 400 million years (Fig. 3-4). The chaotic component consists of the variations in the envelope bounding the eccentricity and inclination curves of Fig. 4. These variations reach several degrees for the inclination, and this mechanism probably explains the current high values of Mercury's eccentricity and inclination.

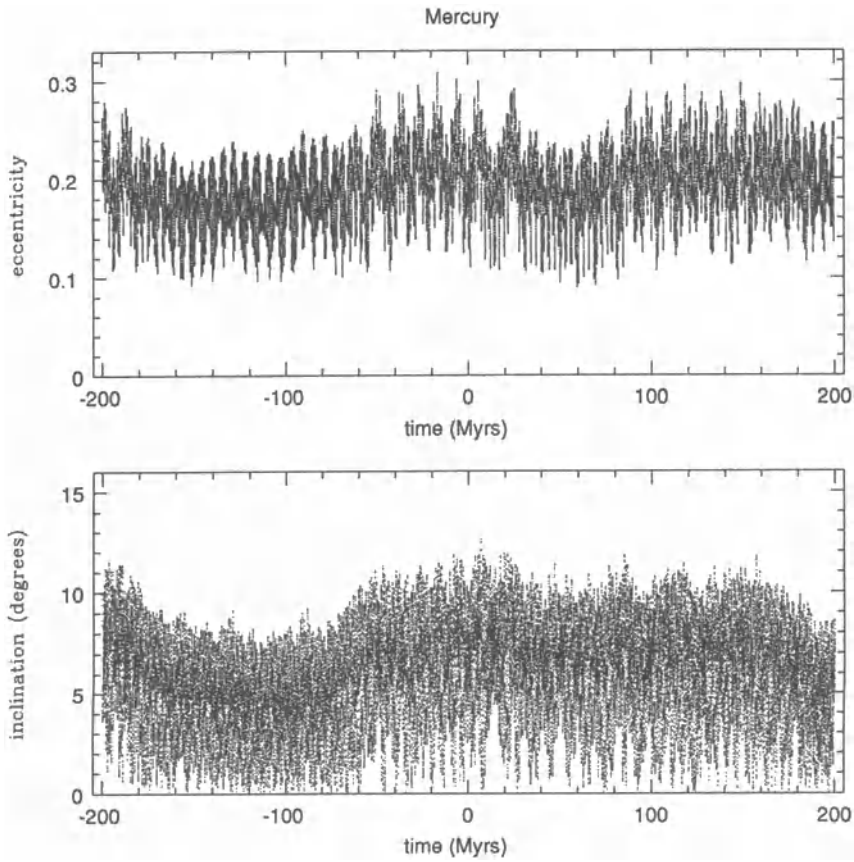


Fig. 4. The chaotic motion of Mercury: Mercury is the planet for which the effect of the chaotic dynamics is the most spectacular. Here are given the computed evolution of the eccentricity (top) and inclination (bottom) of Mercury with respect to time from -200 to $+200$ million years. On each of these curves, two kind of variations can be seen: a rapid variation, with periods of about 100,000 years which correspond basically to the regular part of the solution, as described by Laplace, and a slow variation, which shows the effect of the chaotic dynamics. This variations reaches 0.05 for the eccentricity of Mercury, and several degrees for its inclination. More, the regular variations of Laplace are bounded for an infinite time, but we do not know at the present what are the possible chaotic variations of the eccentricity and inclination of the planets over time span of 5 billion years, comparable to the age and life expectation of the solar system

Instabilities of another sort also manifest themselves in the motion of the solar system's planets. These motions are not present in the orbits, but rather in the orientation of the planets' axes of rotation. Because of their equatorial bulge, the planets are subject to torques arising from the gravitational forces of their satellites and of the Sun. This causes a precessional motion, which in

the Earth's case has a period of about 26,000 years. Moreover, the obliquity of each planet —the angle between the equator and the orbital plane— is not fixed, but suffers perturbations due to the secular motion of the planet's orbit. In particular, for Mars, it was known since the work of Ward (1974) that these perturbations induce large oscillations of about 10 degrees of its obliquity due to the proximity of secular spin orbit resonances. In fact, recent computations showed that the motion of the obliquity of Mars is chaotic (Laskar & Robutel 1993; Touma & Wisdom 1993). Using frequency analysis (Laskar 1990, 1993; Laskar et al.1992b), we could even show that the chaotic zone of the obliquity of Mars extends from 0 to 60 degrees, and such large changes in obliquity were observed over time span shorter than 50 Myr (Laskar & Robutel 1993).

For the Earth, the oscillations of the obliquity, which appear to be the determining factor in the onset of the ice ages, are regular, varying only ± 1.3 degrees about the mean obliquity value of 23.3° (Laskar et al. 1993a). But in the absence of the Moon, the Earth's obliquity would no doubt be chaotic, experiencing strong oscillations ranging from 0° to nearly 85° which would profoundly modify the surface climate (Laskar et al. 1993b). Such behavior was perhaps also experienced by Mercury and Venus in the course of their history, before their rotations were slowed through dissipative effects. In particular, for Venus, this possible instability provides a mechanism for inverting the planet without resorting to large impacts (Laskar & Robutel 1993).

Acknowledgments: It is a pleasure to thank H.S. Dumas for his help.

References

Applegate J.H., Douglas M.R., Gursel Y., Sussman G.J., Wisdom J., 1986, "The solar system for 200 million years", Astron. J. 92, 176–194

Arnold V.I., 1963, "Small denominators and problems of stability of motion in classical celestial mechanics", Russian Math. Surveys 18, 6, 85–193

Bretagnon P., 1974, "Termes à longue période dans le système solaire", Astron. Astrophys. 30, 141-154

Brumberg V.A., Chapront J., 1973, "Construction of a general planetary theory of the first order", Celes. Mech. 8, 335-355

Carpino M., Milani A., Nobili A.M., 1987, "Long-term numerical integrations and synthetic theories for the motion of the outer planets", Astron. Astrophys. 181, 182-194

Duriez L., 1979, "Approche d'une théorie générale planétaire en variables héliocentriques", thèse, Lille

Laskar J., 1984, "Théorie générale planétaire: éléments orbitaux des planètes sur un million d'années", thèse, Paris

Laskar J., 1986, "Secular terms of classical planetary theories using the results of general theory",Astron. Astrophys. 157, 59-70

Laskar J., 1989, "A numerical experiment on the chaotic behaviour of the Solar System", Nature 338, 237–238

Laskar J., 1990, "The chaotic motion of the solar system. A numerical estimate of the size of the chaotic zones", Icarus 88, 266–291

Laskar J., 1992a, "A few points on the stability of the solar system", in Symposium IAU 152, S. Ferraz-Mello ed., 1–16, Kluwer, Dordrecht

Laskar J., 1992b, "La stabilité du Système Solaire", in Chaos et Déterminisme, A. Dahan et al., eds., Seuil, Paris

Laskar J., 1993, "Frequency analysis for multidimensional systems. Global dynamics and diffusion", Physica D 67, 257–281

Laskar J., Quinn T., Tremaine S., 1992a, "Confirmation of Resonant Structure in the Solar System", Icarus 95, 148–152

Laskar J., Froeschlé Cl., Celletti A., 1992b "The Measure of Chaos by the Numerical Analysis of the Fundamental Frequencies. Application to the Standard Mapping", Physica D 56, 253–269

Laskar J., Robutel P., 1993, "The chaotic obliquity of the planets", Nature 361, 608–612

Laskar J., Joutel F., Boudin F., 1993a, "Orbital, precessional, and insolation quantities for the Earth from −20Myr to +10Myr", Astron. Astrophys. 270, 522-533

Laskar J., Joutel F., Robutel P., 1993b, "Stabilization of the Earth's obliquity by the Moon", Nature 361, 615–617

Newhall X. X., Standish E. M., Williams J. G., 1983, "DE102: a numerically integrated ephemeris of the Moon and planets spanning forty-four centuries", Astron. Astrophys. 125, 150-167

Nobili A.M., Milani A., Carpino M., 1989, "Fundamental frequencies and small divisors in the orbits of the outer planets", Astron. Astrophys. 210, 313-336

Quinn T.R., Tremaine S., Duncan M., 1991, "A three million year integration of the Earth's orbit", Astron. J. 101, 2287–2305

Sussman G.J., Wisdom J., 1988, "Numerical evidence that the motion of Pluto is chaotic", Science 241, 433–437

Sussman G.J., Wisdom J., 1992, "Chaotic evolution of the solar system", Science 257, 56–62

Touma J., Wisdom J., 1993, "The chaotic obliquity of Mars", Science 259, 1294–1297

Ward W.R., 1974, "Climatic Variations on Mars: 1.Astronomical Theory of Insolation", J. Geophys. Res. 79, 3375–3386

The One-Dimensional Three-Body Problem: Numerical Simulations

J.L. ROUET [1], R. DUFOUR [2], & M.R. FEIX [2]

[1] UFR Faculté des Sciences, BP 6759, Orléans Cedex 2, France
[2] PMMS/CNRS, 45071 Orléans Cedex 2, France

Abstract. The ergodic properties of a one-dimensional gravitational system belonging to the microcanonical ensemble are studied. This system, constituted of equal-mass particles, exhibits very strong binary structures which prevent the system to be ergodic and then to reach the theoretical curves predicted by Rybicki. The presence of the binary structure (called molecule) is examined through two criterions. The first one is given by a topological property based on the order relation of the one-dimensional systems; the second one is the internal energy of the molecule. At last the study of the molecule stability indicates that it strongly depends on initial conditions.

1. Introduction

One-dimensional systems have been studied since many years. Although unrealistic, they give some interesting ideas on the basic physical concepts. Such is the case for the study of ergodicity for the following reasons:

- The theoretical distribution functions of one-dimension systems have been determined for numerous systems; plasma (Lenard 1961), Boltzmann-gas (Rouet et al. 1993), gravitational systems (Rybicki 1971).
- In the case of some one-dimensional systems it is possible to perform an exact numerical code (Feix 1969), a property which is crucial to test the validity of the ergodic concept.

At last the very-long range field of one-dimensional gravitational particles keeps the system self-trapped and amplifies the formation and live of coherent structures such as binary structures observed in this study.

The validity of the concept of ergodicity has been numerically tested on a very simple system, that is, the Boltzmann gas. Such a system is constituted of point particles, of different masses, moving on a line and experiencing hard-core collisions (for equal masses, the result is trivial). The numerical results fit the theoretical one obtained,

- by supposing the surface of equi-energy uniformly covered (this surface, or hypersurface, is drawn considering the "right" variable $P/\sqrt{m}$),
- because the system has an order relation given by the position of the particles (in this case the particles cannot cross each other and the initial order position relation is preserved during the simulation).

These considerations give two interesting results: we obtain the equi-repartition of the energy whatever the number of particles is, and the position distribution function is independent of the particle masses (cf. Rouet et al. 1993 for more details).

Note that, in this case, for special ratio of particle masses we do not have a filling of the surface of constant total energy because the initial values of the particle velocities are preserved. We verify that the numerical results recover this property which indicates that the round-off errors are not too important, and, at least for this case, do not destroy this conservative set of values.

2. The One-Dimensional Gravitational Three-Body Problem

2.1 Model and Preliminary Results

In the case of a one-dimensional system, the particles are infinite plane sheets of uniform superficial density of mass μ, moving in the perpendicular direction of the sheet. As the field has no divergence at the origin, the particles are allowed to pass freely through each other.

This present study is limited to the case of $N = 3$ particles of the same mass μ.

This system has three invariants which are, the total energy, the total momentum (which can, without any loss of generality, be taken equal to zero), and then the position of the center of mass. Taking into account these integrals of motion, Rybicki derived the position and velocity distribution functions for microcanonical ensemble systems of N particles.

Numerical and theoretical results are compared in Figs. 1 for the velocity and position distribution functions. For this simulation (simulation A) the velocities of all the particles are initially equal to zero and the positions are respectively $x_1 = -15$, $x_2 = 6.5$ and $x_3 = 8.5$ in arbitrary units. The fitting is not good at all and the position distribution function looks like the superposition of two curves of systems containing each two particles.

Other simulations with different initial conditions have been performed but neither give a good agreement with the Rybicki curves.

Consequently it may exists, in addition with the three invariants, another constraint which prevents the system to be ergodic.

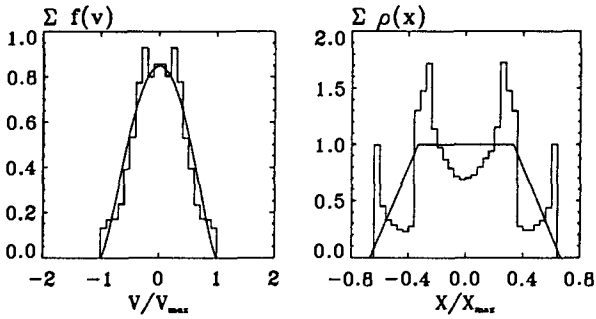


Fig. 1. Sampling at regular time the three positions and velocities obtained with simulation A gives **a:** the velocity distribution function of the three particles, **b:** the position distribution function of the three particles. The theoretical function is drawn with continuous line

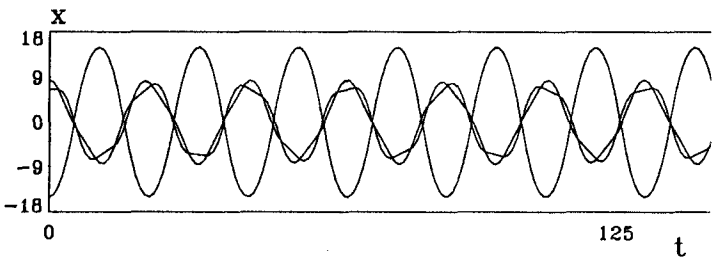


Fig. 2. Trajectories of the three particles for simulation A

2.2 Concept of Molecule

A simple look at the trajectories of the three particles of simulation A helps us to see what happen (see Fig. 2). In fact, two particles make a "couple" which does not break at least after 10^7 crossings. Such a situation is more or less well observed for a large range of initial conditions. The topological order during the evolution of the system gives us an operational diagnostic to know whether a molecule still exists. Suppose molecule $2 \cdot 3$ is formed. Then, the molecule still exists if, particle 1, when entering by one side of the molecule (particle 2 or 3) goes out by the other side (particle 3 or 2). As long as this situation is preserved we will say that molecule $2 \cdot 3$ exists and we will call "topological invariant" this existence. For example, the order of the three particles for simulation A is given in Table 1.

We can also define the internal energy of the molecule which is the energy of the two particles of the molecule in its own frame. This expression reads:

$$U_{23} = \tfrac{1}{4}(v_2 - v_3)^2 + 2\pi|x_2 - x_3| \tag{1}$$

The value of U_{23} is conserved as long as particle 1 is outside the molecule and takes a new value after particle 1 has entered and then left the molecule.

Table 1. Sequence of the relative positions of the three particles for simulation A

3 2 2①①①①3 3 2 3 2 2①①①2 2 3 2 3 3①①①①2 2 3 2
2 3①2 3 2 3①2 3 2 3①2 3 2①3 2 3 2①3 2 3 2①3 2 3
①①3 3 2 3 2 2①①①①3 3 2 3 3①①①①2 2 3 2 3 3①①①

Table 2. Sequence of the relative position of the three particles for simulation B before the topological invariant breaks.

①①3 3 2 2①...①①2 2 3 3①①2 2 3 3①①2 2 3 3①①2①
2 3①2 3①2 ...3 2①3 2①3 2①3 2①3 2①3 2①3 2①3 2①2
3 2 2①①3 3 ...2 3 3①①2 2 3 3①①2 2 3 3①①2 2 3 3 3

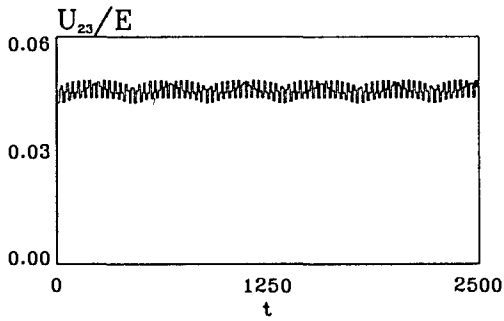


Fig. 3. Ratio of the internal energy divided by the total energy of molecule 2 · 3 for simulation A

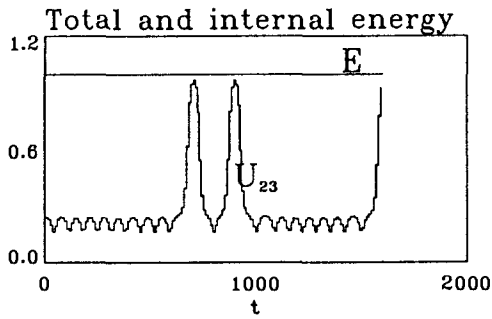


Fig. 4. Internal energy divided by total energy of molecule 2 · 3 for simulation B

Figure 3 gives a small time interval of the variation of U_{23} for simulation A. It oscillates between the two values given by the figure for all the simulation time. Nevertheless, this quiet situation is not a rule and the life of the molecule is sometimes very strange. We start another simulation (simulation B) with $x_1 = -15$, $x_2 = 0.37$, $x_3 = 14.63$ with the three velocities set initially to zero. Table 2 gives the order of the particles and Fig. 4 the internal energy, until the topological invariant breaks, (that is the molecule breaks). At the beginning the internal energy oscillates slightly as for the previous stable molecule of simulation A. Suddenly it increases twice but the topological invariant is not

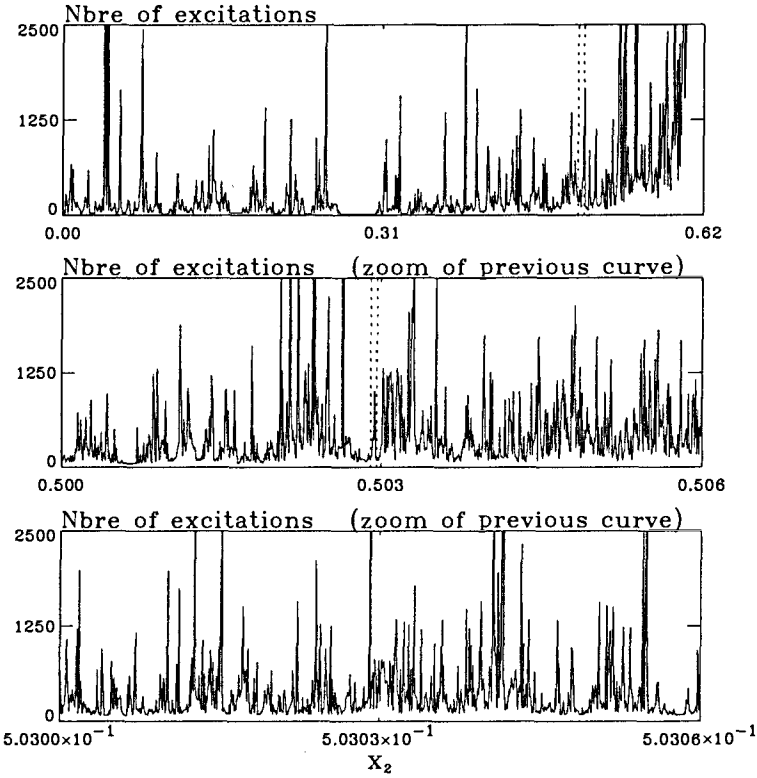


Fig. 5. a: Number of excitations (passage of particle 1 through the molecule $2 \cdot 3$) necessary to destroy the topological invariant as a function of the initial position of particle 2. **b:** Zoom of a small part of curve 5a delimited by the two dotted lines. **c:** Zoom of a small part of curve 5b delimited by the dotted lines

destroyed and the initial values are recovered after a few time. Nevertheless, despite a strong regularity of U_{23}, the molecule breaks suddenly after 146 excitations (passage of particle 1 into molecule $2 \cdot 3$).

The stability of the topological invariant is also a very curious thing. We performed a lot of simulations starting with velocities initially equal to zero and the position of particle 2 taken equal to $x_2 > 0$, particle 1 being initially localized at $x_1 = 15$ and $x_3 = -x_2 - x_3$. This restrictive initial conditions allow us to decide a priori that the molecule is formed by particles 2 and 3.

For each simulation we count the number excitations (passage of particle 1 through the molecule) until the topological invariant breaks (see Fig. 5a). If initial position of particle 2 is bigger than 0.6, the topological invariant is not destroyed after 10^7 excitations and if $x_2 < 0.6$ the number of excitations depends strongly on initial conditions. Zooming a small part of this curve and reiterating once more this process we obtain Figs. 5b and 5c which have more or less the same behaviour, indicating that it has fractal properties.

3. Conclusion

The existence of the very strong binary structure (molecule) prevents the system to be ergodic and consequently the Rybicki curves are not recovered. Unfortunately, the mechanism which keeps such a regularity is not yet understood.

Nevertheless, numerical simulations allow us to imagine processes which, although non physical give useful indications; such is the case of the following experience: as the total energy, which is one of the invariant of the problem, does not depend of the position of the central particle, we can change from time to time this value. Here we relocate particle 2 at random between x_1 and x_3 while v_2 remains the same. This, of course, is a very efficient way to destroy the molecule. In addition, the positions of the three particles are shifted in order not to change the position of the center of mass.

Adding this process to the program, the numerical results fit now the theoretical curves showing that the binary structure effectively is responsible of the previous misfit (cf. Fig. 6).

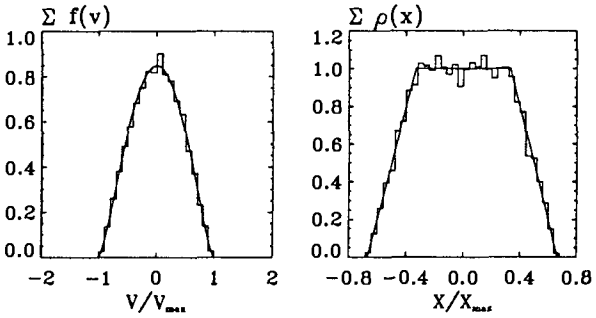


Fig. 6. Velocity and position distribution function when the molecule is destroyed by relocating the central particle at random between the first and the last one. Histogram shape: numerical simulation result. Continuous line: theoretical curve

Finally, we formulate the following hypothesis: the N-body problem is not perfectly ergodic while it exhibits more and more complex structures as N increases. Nevertheless, when N is sufficiently high it will be difficult to determine by computer simulation whether the system is ergodic or not.

Acknowlegdments. The authors thank L. Blanchet for interesting discussions on this topic.

References

Lenard A., 1961, J. Math. Phys. 2 (5), 682
Rouet J.L., Blasco F., Feix M.R., 1993, J. Stat. Phys. 71
Rybicki G.B., 1971, AA & SS 14, 56
Feix M.R., 1969, in: Non-Linear Effects in Plasmas, G. Kalman, M.R. Feix (eds.), Gorden and Breach, New York, p. 151

6. Large-N Limit

Order and Chaos in "Collisionless" Numerical Simulations

Daniel PFENNIGER

Geneva Observatory, University of Geneva, CH-1290 Sauverny, Switzerland

Abstract. The collisionless limit of the N-body problem is often justified by the long 2-body relaxation time. Several other possibilities exist by which stellar systems can evolve and relax faster than previously expected. In particular observations and numerical experiments suggest that fractal distribution functions can spontaneously occur. The understanding of the complex evolution of N-body systems such as galaxies requires presently complementary paradigms.

1. Modelling Stellar Systems

1.1 Epistemological Remark

The scientific process of gaining knowledge about nature (and, more generally, about complex systems, such as chess) always needs the invention of *models* (also called theories), which have the purposes both to reduce nature complexity to a level that at least some human beings can manage, and to select out the most essential aspects of the investigated processes. Therefore models never match exactly nature since some aspects of reality are deliberately discarded, but the abstraction reached in models focuses on the "deeper" (in some sense "truer") properties of nature. On the other hand, by construction modelling will always be a process at the level of some human intelligence. In this view, the possibility of "understanding" parts of nature is no wonder because we select out the aspects of it that we are able to describe in our own words.

Since models truncate reality, scientists have to find out the conditions when the models are falsified, i.e. to specify the limits when the models stop to apply. "Wrong" theories have a narrow, or no scope of applicability, "good" theories a broad one. Absolutely "right" theories can only be invented for systems simple enough to be encompassed by our finite intelligence. Perhaps a complete theory of chess will be invented some day, but the prospects to set up completely right theories of many natural systems are slim due to their enormous complexity.

The Newtonian N-body model of stellar systems is not an exception. Nowadays Newton's laws are no longer taken as absolutely true, yet the N-body model is still useful because it is thought to apply well for long times to real stellar systems. But the complete understanding of the N-body problem in no way would stop the modelling of stellar systems, which are much richer in phenomenons that N point masses. Since the N-body model, expressed by a particular system of differential equations, is already too complex to understand completely, scientists have applied to it the scientific method applied to nature, by making models of the N-body model.

One of these models of the N-body model for large N is the collisionless limit described by the collisionless Boltzmann equation (CBE), which transforms the problem into an integro-differential equation, so using the formal language of mathematics.

Another class of formal systems is provided by various computer models. Numerical computer models have not reached today the esteem that analysis enjoys, in large part because round-off errors introduce some fuzziness in results. Yet computers should not be identified with floating number crunching devices. In many cases computers can manipulate symbols as rigorously as trained mathematicians[1]. In such cases, often with discrete models, computer models can compete in efficiency (with, e.g., differential equation models) for representing complex systems.

So when the divergence between a natural stellar system and an analytical N-body model is as large as the divergence with a discrete computer model, both models should be considered as useful; indeed both models are increasingly wrong at the same rate.

1.2 Historical Remark

Since the pioneer works in the sixties of von Hœrner, Miller, van Albada, Hohl, Hockney, Prendergast and many others, numerical simulations of galaxies have been performed with the aim to reproduce the collisionless limit at which the collisionless Boltzmann equation (CBE) was and is still supposed to apply. Most of these simulations are thought to model the large scale gravitation of galaxies. But the number of particles that can be tackled with present day computers is only about the square root of the number of stars in galaxies, so it has been realised soon that the discrepancy should be investigated, and the unwanted effects of numerical simulations minimised, when possible.

In the early simulations of galaxies, several results have been obtained, but due to the limited amount of independent works and the restricted exploration of the free parameters, it was not clear whether particular results would be general, or whether they would be the consequence of numerical

[1] For example computer algebra may deal with irrational numbers as precisely as human beings do (e.g., it simplifies $\sqrt{2}^2$ by 2, while floating-point number arithmetic may give 1.99999).

artifacts. Therefore the astronomical community did not fully appreciate the generality of some early results. With the years we can stand back and make some kind of synthesis of the past experiments. In particular several N-body results were going against the image people had of galaxies, or the admitted paradigm of the collisionless approximation:

1. Disc galaxy models tend to make a bar, while the general prejudice was that stable disc galaxies should be axisymmetric. This leads to the ideas that evolution can be driven by large scale instabilities (symmetry breaking) much faster than 2-body relaxation, and the most natural states are not always the ones with the highest degree of symmetry.
2. Disc N-body simulations are highly "live", because spiral discs are marginally stable against radial instability. Consequently the assumed steady state of spiral arms is difficult to maintain for many Gyr. More realistically, disc galaxies seems today to constantly generate evanescent spiral arms, and to "swing"-amplify perturbations to large scale. Again this liveliness of galactic diks clashes with a slow evolution of discs driven by 2-body relaxation.
3. N-body particles, like stars in the solar neighbourhood (Wielen 1977; Fuchs et al., this volume), do not conserve well their expected integrals of motion (van Albada 1986). The observational *fact* of stellar diffusion, i.e. increase of oscillation amplitudes about circular orbits as a function of stellar ages with a rate similar to the orbital period, invalidates strongly the collisionless model for the Galaxy. The particle diffusion in numerical experiments, sometimes viewed as a defect of the numerical methods, has the advantage, at least qualitatively, to resemble reality in this aspect.

Therefore time has come to reexamine whether classical assumptions about the collisionless approximation of the N-body problem are effectively relevant. The qualifier "collisionless" is sometimes quoted, since this property is generally more believed than demonstrated to apply. Fortunately, and contrary to galaxies, computer N-body models can be experimented with and the collision properties can be investigated.

2. Difficulties with the Collisionless Approximation

2.1. Existence of Smooth Density or Distribution Functions

It is rarely explicitly stated that the mere *existence* of differentiable mass distributions $\rho(\mathbf{x})$, and even more distribution functions (DF) $f(\mathbf{x}, \mathbf{v})$ is generally *assumed* to apply but not verified to exist in real stellar systems. While the reason of the smoothness of common fluids can be understood intuitively by the fast collisional amplification of perturbations ("molecular chaos"), we lack of such an argument for stellar systems.

Also it is difficult to test whether a smooth density model is relevant in real stellar systems, because, except for the solar neighbourhood, only projections of mass distributions are observed. DF's are also seen in projection, with perhaps only the projection of the first and second velocity moments along the line of sight being accessible. In the case of our Galaxy, we have access to a more detailed information, but the opposite evidences of frequent hierarchical mass distributions are ubiquitous, such as multiple bound stars, groups, clusters, star streams, etc., suggesting rather inhomogeneous models of stellar distributions.

Since stars originate from the cold interstellar medium, it is natural that young stars keep for some time the memory of their initial birth conditions. In the recent years, it has been realised that this cold gas is better described by a fractal model over a range of scales of at least 10^4, with $0.01\,\mathrm{pc} < r < 100\,\mathrm{pc}$ (Scalo 1990; Falgarone 1992; Pfenniger & Combes 1994). A fractal mass distribution in a metric space follows a scaling relation between the mass M and the length-scale r (Mandelbrot 1982):

$$M(r) \sim r^D \,, \tag{1}$$

with D defining the fractal dimension. Observations of the cold interstellar medium suggest $1.4 \lesssim D \lesssim 2$, while a smooth gas would have $D = 3$. For a DF, the same relation holds in the 6D phase space, where r is a distance in phase-space. As soon as D is smaller than the space dimension, in general discontinuities arise in the mass distribution which invalidate a differentiable model.

It is sometimes wrongly believed that an integer D implies a corresponding topological dimension, e.g. $D = 2$ for cold gas in space would mean that the mass distribution is sheet-like. In fact an integer D just describes the scaling of mass with distance, so D constraints very little the smoothness of the distribution.

A theorem (cf. Falconer 1990, Chap. 6) about the fractal dimension of projections of fractal sets states that if the fractal set has a dimension smaller than the projection space, then the projection has in general the same dimension as the fractal set, while if the fractal set has a dimension larger than the projection space, then the projection has in general the dimension of the projection space. Clearly in a stellar system a fractal density implies a non-differentiable DF, which implies that the CBE is not applicable. The converse is not necessarily true. But if the dimension of a fractal DF is larger or equal to 3, then the fractal dimension of the mass distribution is in general 3, so may look relatively smooth. So by observing only mass distributions, and even more so by observing projections of them, we may miss the fact that the DF is not differentiable.

An academic example of an equilibrium self-consistent stellar system with a fractal DF but a smooth density is a smooth disc made of stars on direct and retrograde circular orbits, the fraction of retrograde orbits being distributed

in radius as a Cantor set. Many more examples like this can be invented, showing that among all the possible systems smooth stellar systems form an exceptional class; we know of no compelling reason which would oblige natural systems to belong to the class of differentiable systems.

On the experimental point of view, simulations of "collisionless" particles also suggest that highly inhomogeneous mass distributions occur spontaneously in some cases. For example the Toomre & Kalnajs (1991) shearing sheets experiments represent a small portion of a marginally stable disc. The particles, each one representing a *population* of stars, develop long-range correlations, suggesting a scaling law and a fractal distribution. If more particles would be run at higher resolution, one expects that the scaling relations would then extent over a wider range. In the limit of an infinite N the system would be a mathematical fractal or a multifractal. In such situations the differentiability of the DF is highly questionable since there is no small scale at which a smooth distribution is a good approximation.

Therefore it is perfectly possible that in several cases a better representation of the stellar distribution might be fractal, or even more complicated. But if stellar distributions are sometimes hierarchical the basic assumptions of classical stellar dynamics have to be reexamined. First of all, the tools of calculus such as differential equations are no longer applicable. By simple arguments (Pfenniger & Combes 1994), it is easy to show that in a scale-free hierarchical system the interactions between clusters at the same level of the hierarchy may increase considerably the collisional character of the system. If such a system is virialised at each level, the dynamical time $\tau_{\rm dyn}$ and the physical collision time $\tau_{\rm col}$ scale with a power depending on the fractal dimension D

$$\left.\begin{array}{l}\tau_{\rm dyn} \sim r^{(3-D)/2} \\ \tau_{\rm col} \sim r^{(7-3D)/2}\end{array}\right\} \implies \frac{\tau_{\rm col}}{\tau_{\rm dyn}} \sim r^{2-D}\,. \tag{2}$$

When $D > 2$ the system is collision dominated, and the transition occurs at $D = 2$.

As an illustration let us consider a stellar system such as the optical part of our Galaxy that we suppose made by a uniform distribution of stars. The mass of the star distribution scales as the cube of the size so $D = 3$, but the previous relations so not apply because star subsets are not in virial equilibrium. If we consider the galaxy and the stellar distribution as a one level hierarchical system, the Galaxy being supposed in virial equilibrium, the fractal dimension of the ensemble is then given by

$$D = \log(M_{\rm gal}/M_{\rm star}) \Big/ \log(r_{\rm gal}/r_{\rm star})\,, \tag{3}$$

which gives a dimension $D \approx 1$. Similar considerations for galaxy clusters and super-clusters give $D \approx 1.2$ (Coleman & Pietronero 1992). So in this view of the Galaxy the physical collision time exceeds the dynamical time

by a factor $\tau_{\rm col}/\tau_{\rm dyn} = r_{\rm gal}/r_{\rm star} \sim 10^{11}$. But if our Galaxy would be made purely of molecular clouds with a mass $M_{\rm mc} = 5 \cdot 10^5$ stars, each one with a size $r_{\rm mc} = 30\,{\rm pc}$, the dimension would be $D \approx 2.8$, and the time ratio would drop to $\tau_{\rm col}/\tau_{\rm dyn} \sim 0.025$: the system would be essentially collisional. Thus we see that a hierarchical organisation can change drastically the collisional character of stellar systems. Similar considerations can be made with the 2-body relaxation time.

2.2. Sensitivity of Orbits in Smooth Density Models

Even if the density $\rho(\mathbf{x})$ is differentiable, it does not mean that the stellar system is truly collisionless over time-scales larger than the age of the Universe. We discuss here another instance where the scope of collisionless stellar dynamics is narrowed.

In the original approach of Chandrasekhar (1941, 1943) the relaxation time is not an intrinsic physical effect, it is an estimate of the shortest time-scale after which the collisionless *model* is no longer valid. So in principle, the collisionless approximation can be invalidated if, for *any* reason, the trajectory of a sizable fraction of particles in a smooth mass distribution is shown to deviate significantly from the trajectory they would have in the corresponding N-body representation. Chandrasekhar estimated the relaxation time of a stationary stellar system would be the fastest by 2-body encounters, so he neglected the effects of the mean field, and collective effects.

By considering only a uniform infinite density distribution, the class of potentials is restricted to the exceptional class of integrable potentials. Today we know that generic potentials are non-integrable. They contain resonances and a positive measure of chaotic orbits which amplify exponentially any perturbation of the smooth mean-field model. So integrable systems miss an important feature of generic systems.

In a particular case of a barred galaxy potential the perturbation due to a single star has been found to be able to relax (in the above sense) typical chaotic orbits in a much shorter time than galaxy ages (Pfenniger 1986). In contrast quasi-periodic orbits resist much longer. The orbits in the smooth model deviate from "reality" (N-body model) with a time-scale given by

$$\tau_{\rm orb} = \frac{1}{\lambda_{\rm max}} \ln\left(1 + \frac{\lambda_{\rm max} v}{a}\right) , \tag{4}$$

where $\lambda_{\rm max}$ is the largest Liapunov exponent of the orbit, v the typical stellar velocity, and a the typical amplitude of perturbing accelerations. To determine how long the smooth model is applicable requires the knowledge of the sensitivity of the orbit, given by $\lambda_{\rm max}$, but also of an estimate of the typical level of perturbations a that is present in a real galaxy. In galaxies this level of perturbations is mainly determined by the large "grains", such as the globular clusters, molecular clouds and HI complexes.

Similar conclusions were reached in a subsequent work (Udry & Pfenniger 1988) in which it was shown that the average Kolmogorov entropy ($h = \sum_{\lambda_i>0} \lambda_i$, where λ_i are the 6 Liapunov exponents of an orbit) of a set of orbits in elliptical galaxy models leads to a time-scale $\langle h \rangle^{-1}$ of the order of the galaxy age.

These examples show that the growth of perturbations between the smooth model of a N-body system and the real stellar system is sufficiently fast to invalidate the initial assumptions of long relaxation time. This does not mean that the real system will evolve with the time-scale $\tau_{\rm orb}$, because collective effects are not directly related with the individual orbit stability properties, but it means only that a new limit occurs when considering a collisionless model of stellar systems with non-integrable potentials.

2.3. Phase Mixing in Steady Potentials

If one defines an *orbit* as the subset of phase-space reached over plus and minus an infinite time by the trajectory starting at a particular point, an orbit is by construction a time-invariant subset of phase-space. Thus orbits defined as such are the elementary building blocks of a steady collisionless stellar system.

Suppose we have a stable collisionless stellar system in equilibrium. If a small perturbation is exercised onto it, by definition of stability, we expect that it oscillates around its equilibrium or recovers its steady state with a typical time-scale.

If the collisionless system is purely made of quasi-periodic orbits (tori in phase-space), we expect by the KAM-theorem that a typical small perturbation preserves most of the tori, but it can perturb their uniform density. In order to recover a steady state, every tori should be time independent by being uniformly populated by stars. So the time required to recover a steady state can be estimated by the time to phase-mix a delta function on the tori, *assuming no collective effect.* The same argument can be applied to the other type of orbits with a positive measure, the irregular orbits, which are defined as the orbits which are not periodic or quasi-periodic orbits.

Since arbitrary potentials are generally not integrable, some positive measure of orbits is chaotic. The question is to characterise the mixing rate on different orbit types.

For measuring experimentally the mixing time, we performed numerical computations with non-integrable galactic potentials and the standard map (Pfenniger 1984, 1985). The intrinsic mixing time of orbits is measured by the following procedure:

1. A particular orbit is integrated over a time T, and the fraction of time spend in each cell (called the "occupation" by Schwarzschild, 1979) of a rectangular grid in the configuration or phase spaces is computed.

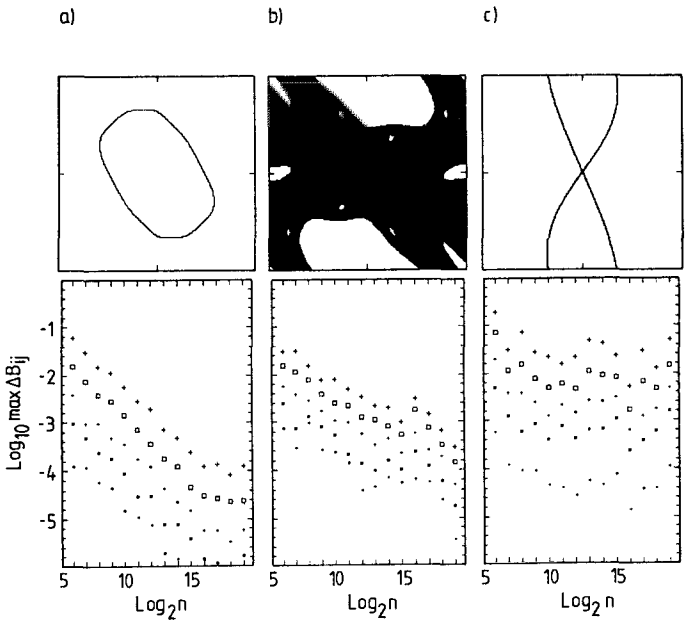


Fig. 1. Fluctuations in the standard map $x' = x + K\sin(x + y)$, $y' = x + y$, as a function of the number of iterations. At the top is shown the final area filled by 2^{19} iterations for a) $K = -1$, $(x_0, y_0) = (0, 2)$ (quasi-periodic orbit), b) $K = 1.8$, $(x_0, y_0) = (0, 0.01)$ (nearly ergodic orbit), c) $K = 0.3$, $(x_0, y_0) = (0, 0.01)$ (weakly ergodic orbit). At the bottom the measure of the fluctuation $\max(\Delta B_{ij})$ is shown for cartesian grids with different resolutions: (large cross) (48×48), (large square) (24×24), (small cross) (12×12), (small square) (6×6), (dot) (3×3)

2. After the time T the same calculation is continued in a second fresh grid for the same time T.
3. The cell occupation in the two grids is then compared. The largest absolute difference of all the cell occupations is taken as a volume independent estimate of the fluctuations.
4. The two grids of occupations are averaged and the result is stored in the first grid. The second grid is zeroed.
5. The calculation is continued for twice the previous total time and the cell occupations are stored in the second grid. The process is repeated at 3., each time doubling the previous integration time, until a fluctuation threshold is reached.

In this way we can measure how fast a particular orbit mixes for a given spatial resolution. The result for grids with lower spatial resolutions can easily be found from the highest resolution grid by averaging the orbit occupations in contiguous grid cells. After having calculated in this way the fluctuations of a large sample of different orbits, a few distinctive results emerge (see Fig. 1):

1. Most quasi-periodic orbits mix fast, the exceptions being obviously the near resonant ones, which can require an arbitrarily long time to cover their torus uniformly.
2. Strongly chaotic orbits with an extended occupation of phase space mix somewhat slower than quasi-periodic ones, principally because they fill a larger volume of phase-space. These orbits jump frequently near any point of the accessible region so the mixing at a finite grid resolution appears as relatively fast.
3. Weakly chaotic orbits can mix *arbitrarily slowly* because their chaotic region is riddled with cantori. Cantori holds orbits in a macroscopic region for arbitrary long times, so when such an orbit escapes into another region, the occupation of phase-space varies strongly.

Therefore weak chaotic motion can invalidate the assumption of steadiness of the collisionless approximation because perturbations on individual phase space invariant blocks (the orbits) are not damped rapidly. This suggests that stable "collisionless" models require not necessarily pure integrable potentials, but, more generally, rapidly mixing orbits.

2.4. Noise in Particle-Mesh N-Body Models

Particle-Mesh N-body models are thought to represent well galaxies because, though N is only about the square root of the number of stars, softening the short range forces decreases the effect of nearby particle encounters. Hohl (1973) has tested that when properly set N-body galaxy simulations do not have mass segregation for more than several Gyr. So for the aspect of mass segregation N-body simulations behave as galaxies are expected to behave.

Yet van Albada (1986) has pointed out that individual particles trajectories of N-body simulations diffuse in action space much faster than expected for a truly collisionless system. This diffusion in action space is actually as fast as the one observed in the Milky-Way for stars born near circular orbits (Wielen's diffusion, 1977). So N-body particles as well as stars do not behave as it is commonly thought they should behave in collisionless stellar systems.

During such Particle-Mesh N-body simulations of disc galaxies (Pfenniger & Friedli 1991) we have tested the following, not described in the paper: during a run of a virialised galaxy model we consider at a chosen point in space separately the potential and forces produced by all the particles within and beyond a radius ε, which defines the "near" and "far" regions of particle interactions. The radius ε is equal to the local softening length, which is of the order of the local grid cell size. After several dynamical times of evolution we Fourier transform the time-series of the forces resulting from respectively the near and far components. The general result is simple: the high-frequency power, the "noise", induced by the far component is typically larger or at least as much important than the high-frequency power from the near component (Fig. 2). High frequencies are defined as the frequencies higher than the one

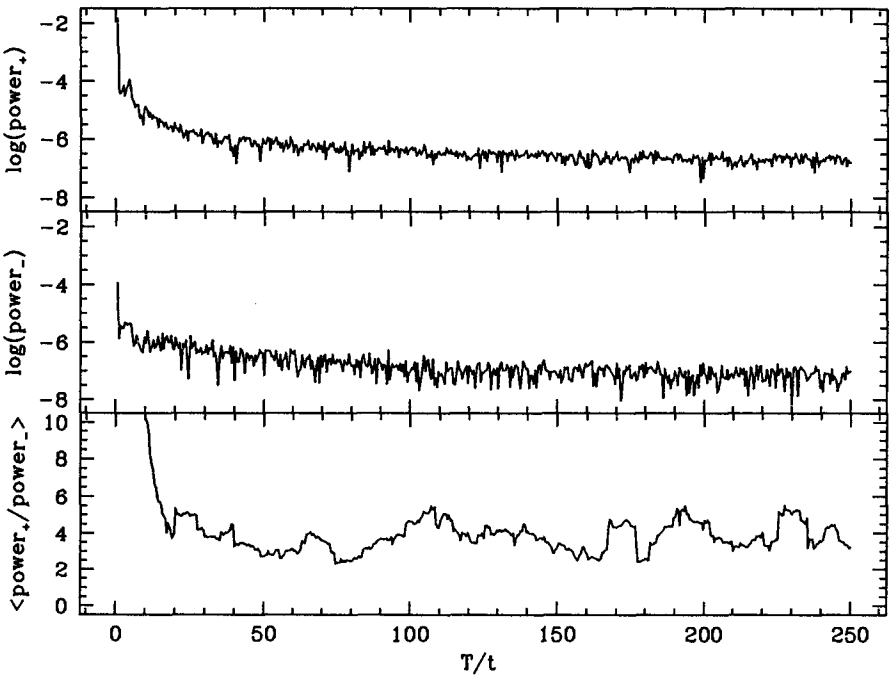


Fig. 2. Power spectrum of the forces of the "far" (top) and "near" (middle) components of a disc galaxy N-body simulation ($N = 2 \cdot 10^5$) at a location in the disc plane ($z = 0$) and at a radius of the order of twice the disc length scale ($R = 7$). The softening radius ϵ is 1.32 there. At the bottom the power ratio is shown averaged over an interval in T/t of 10. The time t belongs to the interval $0.5 \leq t \leq T = 500$. In comparison the dynamical time amounts to ≈ 100. Clearly the far component of the forces has at least as much high-frequency power at the near component

given by the model dynamical time, but lower than the one given by the integration time-step.

This means that the noise reduction achieved by the softening of the nearby interactions has reached a level below the one due to long range interactions. A larger softening of the local forces is not going to improve much the noise level, because the softening softens the near encounters, but not the much more numerous far encounters.

Therefore the collisionless nature of "collisionless" N-body systems is partly illusory because in typical models the fast long range force fluctuations contain more power in high frequencies than the short range ones. The only known way to decrease the force fluctuations due to the far component is, presently, to increase the number of particles. This is a very inefficient way for future progress.

In galaxies softening may be justified on physical grounds. It can be seen as a judicious approximation based on energetics: in galaxies, stellar regions

smaller than the softening length contain much more kinetic energy than self-gravitational energy. At this scale ($\sim$ 100 pc) gravitation plays a little rôle with respect to the kinetic pressure, the local system is automatically in equilibrium and stable with respect to its own gravitation (in other words its Jeans length is larger than the softening length). Therefore short range gravitational interactions can be neglected, justifying the use of grids. With well chosen grids efficient Poisson solvers exist, e.g. using FFT techniques.

3. Order and Chaos in N-body Simulations

We know from observations and from N-body simulations that galaxies come in a restricted variety of shapes, which suggested to Hubble his famous sequence. We have global systematic attractors in these Hamiltonian systems, though the term "attractor" is usually well understood for dissipative systems. When a collisionless ensemble of stars relaxes collectively, if it is slowly rotating it takes either a triaxial $\log \Sigma \approx r^{1/4}$ de Vaucouleurs profile, where Σ is the projected density (van Albada 1982), or if it is rapidly rotating in an initial disc shape, it forms a rotating bar which subsequently rearrange the outer disc profile into an exponential shape (Hohl 1971) similar to the observed galaxy optical discs (Freeman 1970). Why these shapes are preferred is today unclear, but their appearance both in nature and in N-body models suggest a simple underlying principle.

3.1. Stellar Disc Evolution

Typical N-body simulations of equilibrium discs representing galaxies evolve so that the disc forms a rotating bar as large as the turnover radius of the solid body part of the rotation curve, and evanescent spiral arms in the outer disc. After a few rotations, a bar can be unstable with respect to a bending instability *transverse* to the disc plane forming a peanut-shaped bar that looks like a boxy bulge when the galaxy model is seen on edge. Many aspects of these N-body bulges resemble observed peanut-shaped bulges (Friedli & Pfenniger 1990), such as the density profile. The peanut formation by bars was first discovered by Combes & Sanders (1981), and further studied by Combes et al. (1990), Pfenniger & Friedli (1991), and by Raha et al. (1991). This phenomenon will serve as an illustrative example how different existing paradigms can be used for understanding it.

The shape of rotating bars is not random too. In particular the box-peanut shape of bars, that can be called "bulge" depending on the viewing point, is also not explained theoretically. The phase-space structure of these objects possesses a series of singular properties (Pfenniger 1990; Pfenniger & Friedli 1991). For example the Lagrangian points $L_{4,5}$ tend to be marginally stable, the Inner Lindblad Resonance tends to be marginally present, the

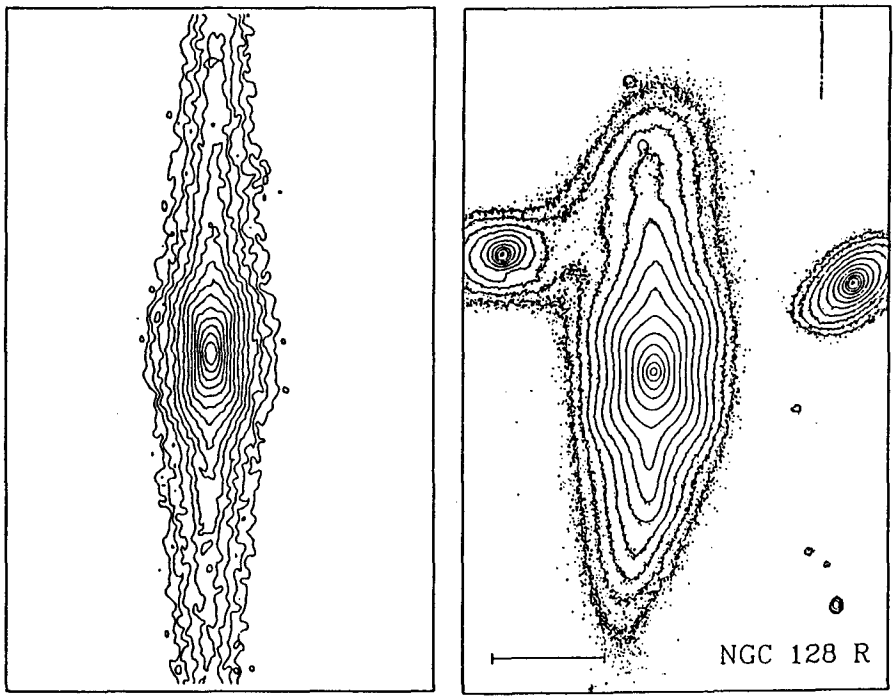


Fig. 3. A N-body model of disc galaxy forms spontaneously a bar, that is subject then to a bending instability perpendicular to the galactic plane. This bending instability saturates with a peanut shape when seen on edge, shown on the left (Combes et al. 1990, Pfenniger & Friedli 1991). The peanut shape can be understood by the existence of 2/2/1 periodic orbits. On the right is shown the peanut-shaped galaxy NGC 128 in the R band (Jarvis 1990). In both frames the isophotes are spaced by half a magnitude (a factor ≈ 1.6)

not populated retrograde family of periodic orbits is vertically unstable over a wide range of the characteristics relevant to the bar, the bump in the peanut-shaped bar occurs at the location (in phase-space) of a 2/1 vertical resonance of the direct periodic orbits.

4. Understanding N-body Simulations

As for natural phenomenons, after the mere description of N-body runs we need more explanations for improving our understanding of them. Below we describe two paradigms that have been applied in the past years.

4.1 Paradigm a): the Linearised CBE

The peanut-shaped bar formation is explained by Raha et al. (1991), Sellwood (1992) with the help of the CBE paradigm applied to the simpler situation of an infinite homogeneous and non-rotating sheet. It can be shown that when the velocity dispersion parallel to the sheet is about ten times the one perpendicular to the sheet, in the linear regime a bending instability is triggered; it is called the "fire-hose" instability (see e.g. Fridman & Polyachenko 1984).

In contrast the N-body model is inhomogeneous, finite, rotating, and from our experiments it does not need to fulfill the fire-hose instability criterion in order to bend.

Since the bending instability goes rapidly beyond the linear regime, the fire-hose instability analogy can not help for understanding the subsequent evolution. The justification to identify the N-body run behaviour and the infinite slab one as similar relies basically on the knowledge of the type of particle motion, i.e. *on the orbits*: in both cases we have elongated orbits along the plane.

Ideally, for applying this approach to the N-body collisionless model, one should have a self-consistent DF on which one could calculate the growing modes with the coupled CBE and Poisson's equation. The advantage would be that the effect of self-gravity would be included and a broad class of instabilities could be investigated (yet the N-body exponential instability is suppressed by the collisionless hypothesis). But even in this ideal case, only the linear regime can be reasonably described. The final explanation is given in term of the key-word "mode": such and such instability occurs because a corresponding growing mode exists.

In summary the CBE is in principle a very useful tool able to describe many aspects of N-body collisionless models. Yet the difficulties of its applications are formidable, and in any case, the solution of the linearised CBE-Poisson equation is always an incomplete description.

4.2 Paradigm b): Periodic Orbits

Poincaré (1892), when considering the general problem of solving sets of differential equations such as the N-body problem understood that the richness of the solutions was much too broad to be fully described by a finite number of terms. Therefore he proposed a strategy for describing phase space: first enumerate the fixed points, and then the periodic orbits (PO). Around the stable periodic orbits other orbits have similar topological properties, so their description is a way to *summarise* the complexity of dynamical systems. The mental economy made offers a partial "explanation" à la Mach.

The generality of the main PO's in galactic potentials is insufficiently appreciated. The main PO *shapes* depend only on the symmetries of the potential: for example in Stäckel potentials the shape of the orbits is independent on the particular Stäckel potential since they can be found knowing

only the particular choice of ellipsoidal coordinates (see de Zeeuw 1985). As corollary the same applies in cartesian, spheroidal and spherical coordinates.

The robustness of PO's is also to mention. When the potential of a galaxy model is perturbed with low spatial frequencies, the main PO's are hardly changed, contrary to the density which depends on the second derivatives of the potential. Thus for studying the PO's in galactic potentials it is not necessary to know precisely the mass distribution, but the main symmetries of the galaxy.

But for high spatial frequency perturbations of the potential exactly along an orbit this is no longer true, a small perturbation may strongly perturb the orbit stability or the orbit shape (Martinet & Pfenniger 1987).

For weak dissipative perturbations, stable PO's and fixed points are also robust in the sense that the trajectories diverge from the unperturbed trajectories much slower than other orbits (Pfenniger & Norman 1990). They can even become attractors in the dissipative regime.

So in general PO's and fixed points are the main structures to look at when considering possible perturbations to a stable stellar system. If the perturbation is a low frequency change of the potential, one can be confident that the effect on the main PO's will be small. On the contrary if a high-frequency perturbation acts precisely on the path of important PO's, a drastic change in the stellar system may result.

For example the keystone of a triaxial stellar system regular near the centre is the origin, because the main PO's, the axial orbits, can be destabilized by the accumulation of a small mass near the centre.

Another example results from the knowledge of the PO's in a rotating bar potential (Contopoulos & Papayannopoulos 1980, Athanassoula et al. 1983). A rotating bar is possible because in a well defined range of conditions major PO's elongated along the bar exist. But when the central density increases, these elongated orbits are replaced by PO's perpendicular to the bar. Beyond a critical central mass concentration no self-consistent bar is possible, and a real bar has to dissolve. This is an important process of galaxy evolution leading to understand the transformation of barred galaxies into non-barred ones, and the possible build-up of bulges after the disc formation (Hasan & Norman 1990; Pfenniger & Norman 1990; Hasan, Pfenniger & Norman 1993).

In the case of the orbital structure of the peanut-shaped bars, we have shown that the bending instability is intimately linked with a 2/1 vertical resonance of the elongated PO's supporting the bar (Pfenniger & Friedli 1991). This orbital resonance exists before and after the growth of the bending instability, and its location coincides with the one of the instability. The bifurcating orbits associated with the resonances trap the non-linear saturation phase of the bending instability, explaining the final peanut shape.

The advantages of using PO's as paradigms for understanding complex stellar systems are: accurate potential models are not required, the PO's are

independent of a particular DF and its differentiability, the integrability of the potential is not necessary.

The disadvantages are that self-gravity effects are not included, and the self-consistency of the potential is not guaranteed (not any potential can be a solution of Poisson's equation since gravitational charges must be positive).

4.3 Summary

We don't have a strong enough paradigm to describe the full complexity of stellar systems such as galaxies. We have seen in the particular case of the bending instability of bars that we need more than paradigms a) and b) to understand galaxies in detail. For the present time paradigm a) and b) are necessary and complementary.

5. Conclusions

The full understanding of the N-body problem in no way would stop our quest of explanations about stellar systems. N-body systems are rich in instabilities at all scales, from the "microscopic" exponential instability of the N-body problem much discussed at this conference, to the macroscopic bar instability. Each instability is a new source of chaos, and an opportunity for the system to evolve. Often models break down when the sensitivity to perturbations is exacerbated by instabilities. We need to enlarge the concept of "relaxation time" by the more general description of the various evolution time-scales.

References

Athanassoula E., Bienaymé O., Martinet L., Pfenniger D., 1983, A&A 127, 349
Chandrasekhar S., 1941, ApJ 93, 285 and 323
Chandrasekhar S., 1943, ApJ 97, 255
Coleman P.H., Pietronero L., 1992, Phys. Rep. 213
Combes F., Debbasch F., Friedli D., Pfenniger D., 1990, A&A 233, 82
Combes F., Sanders R.H., 1981, A&A 96, 164
Contopoulos G., Papayannopoulos Th., 1980, A&A 92, 33
de Zeeuw T., 1985, MNRAS 216, 273 and 599
Falconer K.J., 1990, Fractal Geometry, Wiley, Chichester
Falgarone E., 1992, in: Astrochemistry of Cosmic Phenomena, P.D. Singh (ed.), Kluwer, Dordrecht, p. 159
Freeman K.C., 1970, ApJ 160, 811
Fridman A.M., Polyachenko V.L., 1984,Physics of Gravitating Systems, Springer, New-York
Friedli, D., Pfenniger, D., 1990, in: Bulges of Galaxies, ESO Conf. and Workshop Proc. 35, B.J. Jarvis, D.M. Terndrup (eds.), p. 265
Hasan H., Norman C.A., 1990, ApJ 361, 69
Hasan H., Pfenniger D., Norman C.A., 1993, ApJ 409, 91

Hohl F., 1971, ApJ 168, 343
Hohl F., 1973, ApJ 184, 353
Jarvis B.J., 1990, in: Dynamics and Interactions of Galaxies, R. Wielen (ed.), Springer, Berlin, p. 416
Mandelbrot B., 1982, The Fractal Geometry of Nature, Freeman, San Fransisco
Martinet L., Pfenniger D., 1987, A&A 173, 81
Pfenniger D., 1984, A&A 141, 171
Pfenniger D., 1985, Sur la Dynamique Stellaire des Galaxies Barrées, Thesis No. 2156, University of Geneva
Pfenniger D., 1986, A&A 165, 74
Pfenniger D., 1990, A&A 230, 55
Pfenniger D., Combes F., 1994, A&A, in press
Pfenniger D., Friedli D., 1991, A&A 252, 75
Pfenniger D., Norman C., 1990, ApJ 363, 391
Poincaré H., 1892, Les méthodes nouvelles de la mécanique céleste, Vols. 1–3, Paris, Gauthier-Villars, 1957, New-York, Dover
Raha N., Sellwood J.A., James R.A., Kahn F.D., 1991, Nat 352, 411
Scalo J., 1990, in: Physical Processes in Fragmentation and Star Formation, R. Capuzzo-Dolcetta et al. (eds.), Kluwer, Dordrecht, p. 151
Sellwood J.A., 1992, in: Evolution of Interstellar Matter and Dynamics of Galaxies, J. Palouš, W.B. Burton, P.O. Lindblad (eds.), Cambridge University Press, Cambridge, 343
Schwarzschild M., 1979, ApJ 232, 236
Toomre A., Kalnajs A.J., 1991, in: Dynamics of Disc Galaxies, B. Sundelius (ed.), Göteborg University, p. 341
Udry S., Pfenniger D., 1988, A&A 198, 135
van Albada T.S., 1982, MNRAS 201, 939
van Albada T.S., 1986, in: The Use of Supercomputers in Stellar Dynamics, P. Hut, S. McMillan (eds.), Springer, Berlin, p. 23
Wielen R., 1977, A&A 60, 263

On the Permissible Percentage of Chaotic Orbits in Various Morphological Types of Galaxies

L. MARTINET

Geneva Observatory, CH-1290 Sauverny, Switzerland

Abstract. During the recent years, a lot of papers have dealt with the triggering and growth of chaotic orbital behaviour in dynamical models of galaxies. A question remains still open: Which is the tolerable percentage of chaotic orbits in self-gravitating equilibrium models describing the various morphologies of observed galaxies? Selected results will be presented as a step towards the understanding of this question. They concern triaxial systems such as barred and elliptical galaxies.

1. Introduction

Orbital approaches are complementary to other ways of analysing the structure and the evolution of galaxies. They provide detailed information about dynamical phenomena taking place within them such as the existence of robust stable periodic orbits and the amount of matter trapped around them, or still different kinds of instabilities able to trigger chaotic behaviours. Inferences on self-gravitation and global evolutionary aspects are not decisive from such approaches, which are nevertheless useful as prerequisite to the construction of self-gravitating models.

Roughly spoken, a system is called ergodic if any trajectory (except for a set of null measure) fills densely its energy surface. The sequence of points in a space of section, corresponding to such an ergodic trajectory would fill densely the space of section. In spite of the widely-held use of the term of "ergodicity" in works dealing with galactic dynamics problems, ergodic orbits strictly as such, are not found in smooth "realistic" as well as in noisy models of galaxies. Apart from very peculiar cases (see below), at the very most we may speak about chaotic behaviour in *some* regions of phase space. Some other adjectives are also used in the literature: irregular, wild, erratic, semi-ergodic... This behaviour implies that in a space of section the sequence of points corresponding to the trajectory jumps more or less randomly in the fraction of space left open by the invariant curves which correspond in contrast to a regular behaviour.

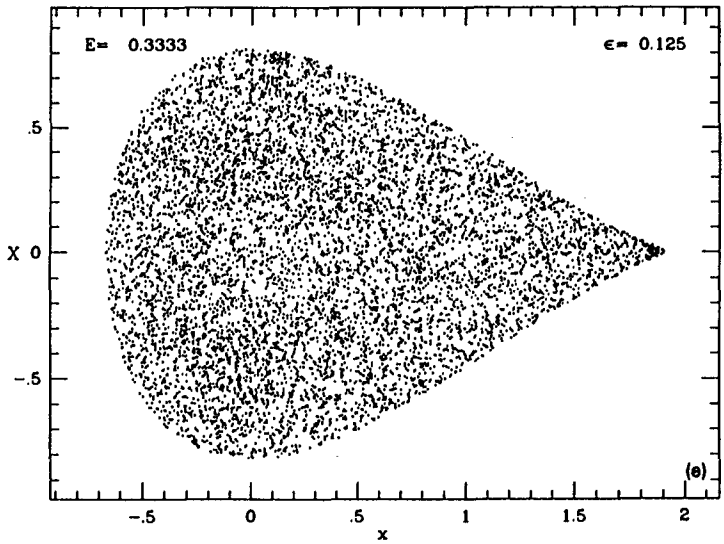


Fig. 1. Surface of section (x, X) for particular initial conditions in the Hamiltonian H_{34} described in the text

In fact, as often as not, a chaotic behaviour is characterised by the presence of *cantori* which, by any means, are not to be confused with ergodicity.

2. A Very Peculiar Case of "Ergodic" Orbital Behaviour

In a study of the transition between the 3-particle Toda lattice (a typical integrable dynamical system) and the Hénon-Heiles model (1964) which displays a typical semi-ergodic behaviour, Udry & Martinet (1990) have considered the particular Hamiltonian

$$H_{34} = H_3 + \varepsilon H_4 \tag{1}$$

where H_3 and H_4 respectively correspond to the 3^{rd} and 4^{th} order expansion term of the Toda lattice. H_3 is the Hénon-Heiles Hamiltonian.

The particular critical case $\varepsilon = 1/8$ is interesting. It separates the cases of bounded motions ($\varepsilon > 1/8$) and unbounded motions ($\varepsilon < 1/8$) in the (E, x) plane. At $E = 1/3$, which is the escape energy for the potential, the limiting curve $E(x)$ of the accessible region has an inflexion point ($\mathrm{d}E/\mathrm{d}x = 0$ and $\mathrm{d}^2E/\mathrm{d}x^2 = 0$). For such conditions, the surface of section $(x, \dot{x})$ is apparently densely covered by randomly distributed consequents (Fig. 1). As far as I know, such ergodic behaviour has never been observed in orbital studies of galaxies.

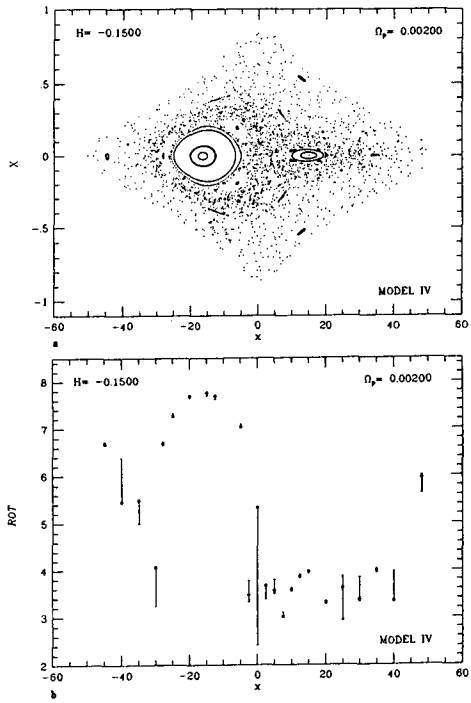


Fig. 2. a) Surface of section, and b) rotation number curve rot(x) for a particular value of the Hamiltonian in a triaxial model with axis ratios (1 : 0.25 : 0.125)

3. Interaction of Resonances in Galaxies, Heteroclinic Orbits

Gerhard (1985) looked for the perturbations of integrable systems which are consistent with observations of early-type galaxies as well as approximately preserving the regular orbital structure of integrable potentials. Limiting his treatment to small perturbations and to homoclinic orbits in order to be able to use the Melnikov integral technique, the author finds that only perturbations such as $\cos m\phi$ ($m = 0, 2, 4$) modest ellipticity gradients or small figure rotation are working in this context. However, it seems to be clear that in realistic systems, a chaotic behaviour can be amplified by resonance interactions and the presence of heteroclinic orbits. Such a situation was described in the inner regions of an axisymmetric model of our Galaxy for stars with small angular momentum (Martinet 1974).

Evidence for chaos, triggered by resonance interactions in triaxial models of galaxies, has been given by Martinet & Udry (1990) in connection with the adopted morphology for these systems. We studied systematically orbits in slowly rotating modified Hubble profile models of various axis ratios ($a : b : c$): I) a nearby spherical one, II) a Schwarzschild model (1 : 0.625 : 0.5), III) a strongly triaxial one and IV) a bar. Surfaces of section as well as the rotation number "rot" attached to invariant curves have been obtained for

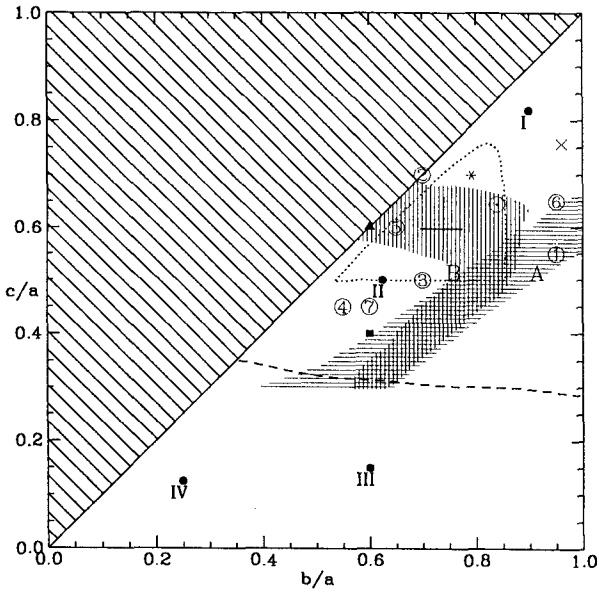


Fig. 3. Axis ratio diagram showing
a) the predicted position of real systems inferred from observational data (various symbols),
b) N-body equilibrium models ($1-7$),
c) modified Hubble profile analytical models (roman numbers)

different values of the Hamiltonian. An example is given in Fig. 2 for model IV. Rational values of "rot" correspond to resonant periodic orbits. When the orbits are not regular, it is impossible to define "rot" (the invariant curves are dissolved) without ambiguity (bottom figure) and discontinuities in "rot"(x) appear, indicating the range of resonances in interaction responsible of chaotic behaviour apparent in the surface of section (top figure).

The main result of this investigation is that only models I and II show moderate chaotic regions (cantori) with not really detected resonance interactions, possibly confined to a very narrow range of rational numbers. On the contrary, models III and IV develop important chaotic regions: for model IV for example, resonances in the range $1/5 <$ "rot" $< 2/3$ are implied in the set-up of chaos. Figure 3 shows that in an existence diagram of axis ratios, our models I and II are located in the region occupied by real triaxial galaxies or bulges (various symbols) according to predictions inferred from observations. On the contrary, the highly triaxial models III and IV are in a region devoided of such real systems. This could be a sketch of constraint on the possible shape of triaxial galaxies. Too highly triaxial objects might not exist because real equilibrium systems would not tolerate too much chaos! In this context, we may mention that either N-body end-products of non-dissipative collapses (Udry 1993) or simulated elliptical systems resulting from mergers (Barnes 1989), have axis ratios consistent with the observational predictions as well as with our models I and II.

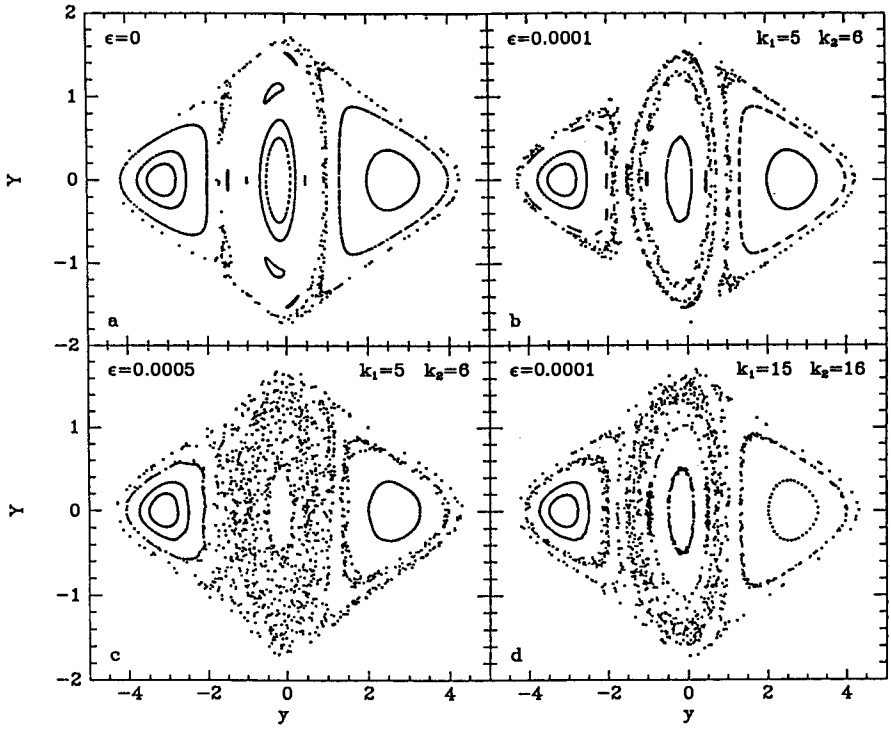


Fig. 4. Effect of a high-frequency sinusoidal potential perturbation on the orbital behaviour in the modified Hubble profile model. See text for the meaning of the parameters

4. Effects of Asymmetries or Noise on the Orbital Behaviour in Triaxial Systems

The gravitational attraction of a co-rotating nearby body is able to deform the shape of orbits in the given systems defined above. In particular, instead of having a main stable orbit x_1 which bifurcates into two stable branches (e.g. Martinet & de Zeeuw, 1988), the asymmetry triggered by an eccentric Plummer sphere leads to a continuous deviation of x_1 from the plane, gradually becoming a banana-shaped orbit (Udry 1991). We will come back below to the problems rising from the existence of this kind of centrophobic orbits. The addition of a high frequency sinusoidal function to the given potential to represent locally some noise modifies the isodensity contours. Udry suggested the following form

$$\Phi = \Phi_0 \left[1 + \varepsilon \sin\left(\sum_i k_i x_i \right) \right] \tag{2}$$

for the noisy 3-D potential. Φ_0 is the potential defined in the previous section. As shown in Fig. 4 displaying surface of section for several cases of the amplitude ε and the frequencies k_i of the noise in the 2-D case chaotic behaviour is favoured by such perturbations. As ε and (or) k_i increase, the invariant curves become thicker and are then progressively destroyed.

5. Barred Galaxies

The topic of dynamics in barred galaxies has recently been extensively reviewed by Sellwood & Wilkinson (1992). Here we summarise again some important conclusions inferred from recent works about the relation between the shape of bars and the onset of chaos. The essential point is that if the axis ratio of the bar in the plane of the disc is larger than 3 to 4, and/or the mass of the bar is larger than 1/4 of the total mass inside corotation, an extended chaotic region can occur in the inner parts of the galaxy and that the corresponding orbits cannot enhance the bar anymore. The stability of the main periodic orbit x_1 along the bar is necessary to maintain the barred structure (for details of the 2-D case see e.g. Athanassoula et al. 1983). These results suggest that models with the properties just mentioned should be excluded for a self-gravitating barred galaxy. As mentioned in the introduction, such predictions taking account collective effects need to be confirmed.

The effect of a compact mass at the centre of a galaxy has been reported by Hasan & Norman (1990, and references therein). The percentage of phase space volume occupied by direct orbits along the bar decreases from 50% in absence of such a compact mass or if the axis ratio a/b of the bar is ~ 2 to $10-15\%$, for instance, if the compact mass is one tenth of the total mass or if $a/b \sim 4$. At the present time, the only equilibrium model existing for a 2-D barred galaxy is the numerical one described by Pfenniger (1984b). In this work it is shown that the percentage of semi-ergodic stars may be as large as 30% but more probably below 10% if the axis ratio is 4 and the mass of the bar is 1/5 of the total mass. Locally, however, around the Lagrangian points at the end of the bar, they can be 100% of semi-ergodic stars.

For a 3-D barred galaxy the estimation of percentages of chaotic orbits is more complicated. From Pfenniger (1984a), it appears that semi-ergodicity is favoured by instability strips perpendicular to the galactic plane. The bar growth is limited by not too large axis ratios a/b and a/c. A too thin bar ($a/c \sim 10$) induces a lot of chaos. In a 3-D strongly barred galaxy, the resonant family 4 : 4 : 1 has been proved to have a complex unstable part (Pfenniger 1985). The resulting orbital diffusion could be a possibility to populate the inner halo.

Much work is necessary to estimate the permissible percentage of chaotic orbits in this case. 3-D equilibrium models do not exist at present time.

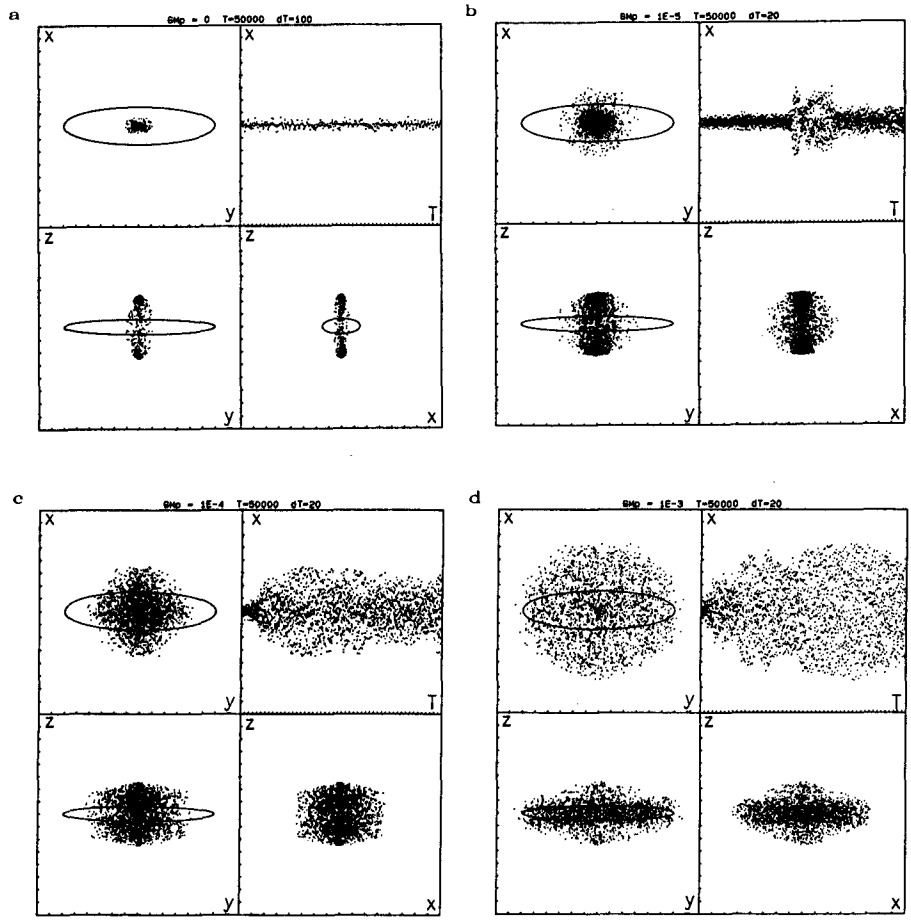


Fig. 5. Diffusion in various projection planes of stochastic orbits starting close to the z-axis orbits in a galactic potential with a central mass GM_p. At the upper right of each frame, the x-coordinates versus time are plotted

6. Complex Instability in Triaxial Systems

This is a new phenomenon which appears in dynamical systems with 3 (or more) degrees of freedom: the Jacobian matrix, associated with the linearised transformation describing the motion close to a periodic orbit has all its eigenvalues complex and outside the unit circle. Real eigenvectors do not exist. It is yet not quite clear in which situations complex instability would produce a great amount of chaotic motions. Important zones of complex instabilities have been found in rather academic potentials (Contopoulos & Magnenat 1985). We mention here an interesting case which could play a role in the equilibrium and stability of real galaxies. In modified Hubble profile

models as well as in other realistic models such as Schwarzschild's potential (Heissler et al. 1982) it has been established that, in absence of rotation or of central compact mass, the (shortest) z-axis orbits are stable, then unstable, finally double-unstable for more and more energetic orbits. The change of stability occurs at bifurcation points with stable and unstable anomalous (inclined) orbits. When we introduce rotation or a compact mass (Martinet & Pfenniger 1987), the z-axis may become partially complex unstable. We observed the case of a compact mass larger than one-thousanth of the total mass which makes the z-axis complex unstable from $z = 0$! The effect of such an instability is described in Fig. 5: the importance of stochastic diffusion of orbits starting near the z-axis is increasing with the central Plummer sphere mass. The diffusion time by this process could be shorter than the Hubble time if the central mass makes most of the z-axis unstable.

7. Boxlets

For a wide class of non-rotating triaxial potentials, most of the phase space is occupied by four well known major families: box, inner and outer long axis tubes, short axis tubes. The x-axis periodic family is generic for the boxes provided that it is stable. Boxes appear essential for the construction of equilibrium models of triaxial systems. However, if there is a sharp variation of the density profile near the centre, the x-axis orbits can be unstable. Centrophilic boxes could be replaced by centrophobic boxlets such as bananas (Miralda & Schwarzschild 1989).

Two points on boxlets need further investigations: firstly, the fact that this shape is not always compatible with the shape of the imposed density. This may cause a problem for the construction for equilibrium models. Secondly, if the boxlets are unstable, they can participate to diffusion of chaotic behaviour. The question concerning what percentage of such orbits is permissible for the construction of self-gravitating systems is still open. This question is still more complicated for rotating systems (Martinet & Udry 1990).

8. Concluding Remarks

- Existing morphologies of galaxies seem to be incompatible with too high percentages of semi-ergodic orbits. Slowly rotating triaxial dynamical models with a major to minor axis a/c ratio larger than 2.5 or fast rotating barred systems with an axis ratio a/b larger than 3 or 4 display important chaotic behaviours. Systems with such morphological features are in fact apparently not observed.
- N-body equilibrium figures obtained by gravitational collapse of initial rotating, anisotropic bodies have b/a from 0.5 to 1 and $c/a \gtrsim 0.45$ in

agreement with predictions of observations. It is the same for Barnes mergers ($a : b : c = 1 : 0.95 : 0.65$).
- N-body bars in discs have axis ratios $a/b \lesssim 4$.
- Quantitative estimations of percentages of semi-ergodic orbits in 3-D systems compatible with observed morphologies is an open question.
- The response to a given potential is not sufficient for answering the initial question. Furthermore galaxies evolve, bars may grow, then dissolve. Ellipticals may accrete. Gas and its various interactions with stars (star formation, gaseous response etc.) contribute to modify the structure of the systems. Does the secular evolution lead such systems to a state close to an integrable potential with modest percentage of chaotic orbits, as suggested by Gerhard (1985)?

References

Athanassoula E., Bienaymé O., Martinet L., Pfenniger D., 1983, A&A 127, 349
Barnes J., 1989, Nature 338, 123
Contopoulos G., Magnenat P., 1985, Celest. Mech. 37, 387
Gerhard O.E., 1985, A&A 151, 278
Hasan H., Norman C., 1990, ApJ 361, 69
Hénon M., Heiles C., 1964, AJ 69, 73
Heissler J., Merritt D.R., Schwarzschild M., 1982, ApJ 258, 490
Martinet L., 1974, A&A 32, 329
Martinet L., Pfenniger D., 1987, A&A 173, 81
Martinet L., Udry S., 1990, A&A 235, 69
Martinet L., de Zeeuw T., 1988, A&A 206, 269
Miralda J., Schwarzschild M., 1989, ApJ 339, 752
Pfenniger D., 1984a, A&A 134, 384
Pfenniger D., 1984b, A&A 141, 171
Pfenniger D., 1985, A&A 150, 112
Sellwood J., Wilkinson A., 1993, Rep. Prog. Phys. 56, 173
Udry S., 1992, A&A 245, 99
Udry S., 1993, A&A 268, 35
Udry S., Martinet L., 1990, Physica D 44, 61

Minimum Energy States of a Self-Gravitating System

J.-J. ALY

Service d'Astrophysique - CE Saclay - 91191 Gif sur Yvette Cédex - France

Abstract. We prove the existence of a unique global minimum energy state for a self-gravitating system whose mass and a "quasi-entropy" assume a priori given values.

1. Introduction

Many authors interested either in the stability of the equilibria of self-gravitating systems (e.g. Ipser 1974; Ipser et al. 1979), or in their "violent relaxation" to an equilibrium (e.g. Tremaine et al. 1986; Wiechen et al. 1988; Aly 1989), have discussed the problem of the existence and determination of their minimum energy states under some constraints. In this paper, we reconsider this problem in the case the mass and a "quasi-entropy" (Tremaine et al. 1986) take a priori prescribed values. Our main emphasis is on the existence and properties of *global energy minimizers.*

2. Statement of the Problem

Consider a self-gravitating system whose state can be described by a distribution function $f(x,v) = f(\xi)$. f is a nonnegative function defined over the phase-space R^6, with $f\mathrm{d}\xi$ representing the amount of mass contained in the volume element $\mathrm{d}\xi$. The mass of the system is thus given by

$$M[f] = \int f(\xi)\, d\xi \,, \tag{1}$$

and its energy by (G denotes the gravitational constant)

$$E[f] = \tfrac{1}{2}\int |v|^2 f(\xi)\,\mathrm{d}\xi - \tfrac{1}{2}G\int \frac{f(\xi)f(\xi')}{|x-x'|}\,\mathrm{d}\xi\mathrm{d}\xi' \,. \tag{2}$$

To f, we associate the "quasi-entropy" (Tremaine et al. 1986)

$$S[f] := -\int C(f)\,\mathrm{d}\xi, \tag{3}$$

where $C : R^+ \to R^+$ is a given differentiable convex function which vanishes at the origin ($C(0) = 0$) and satisfies the technical condition

$$C(f) \geq cf^p \qquad \text{and} \qquad C(f) =_{f\to\infty} O(f^p) \tag{4}$$

for some constants $c > 0$ and $p > 1$.

The problem to be discussed here is the following one: let $M_0 > 0$ and $S_0 < 0$ be given numbers, and $\mathcal{H}$ be the set

$$\mathcal{H} := \{f | f : R^6 \to R^+; M[f] = M_0; S[f] = S_0; |E[f]| < \infty\}. \tag{5}$$

Then *is there in $\mathcal{H}$ a function f^- which globally minimizes the energy?*

3. Lower Bound on the Energy

Clearly, for our problem to have a solution, the energy needs to be bounded from below in $\mathcal{H}$. That this is the case depends on the value of p:

- If $1 < p < 9/7$, then

$$E^- := \inf_{\mathcal{H}} E[f] = -\infty. \tag{6}$$

  It is always possible to find in $\mathcal{H}$ an f with an energy as negative as we want, and the minimization problem has no solution.
- If $9/7 \leq p$, then the inequality (Aly 1989)

$$-\infty < E^* := -kM_0^{(7p-9)/3(p-1)} |S_0/c|^{2/3(p-1)} \leq E^- := \inf_{\mathcal{H}} E[f] \tag{7}$$

  holds, whith $k := (2/5)^5 (42\pi)^{2/3} G^2$. Thus the energy is bounded from below on $\mathcal{H}$, and a minimizer f^- may possibly exist.

It is worth noticing that the inequality $E^* \leq E^-$ in Eq. (7) reduces to an equality in the case where $C(f) = c|f|^{9/7}$, the infimum E^- being actually reached by the *Plummer's distribution function*

$$f^-(x,v) = K\left(-\frac{|v|^2}{2} + \frac{GM_0}{(a^2+|x|^2)^{1/2}}\right)^{7/2} \theta\left(-\frac{|v|^2}{2} + \frac{GM_0}{(a^2+|x|^2)^{1/2}}\right), \tag{8}$$

where K and a are positive constants which can be computed from the values of M_0 and S_0, and θ is the usual step function.

4. Existence and Properties of Minimizers

When $p \geq 9/7$, it can be proven that:

- Whichever be the values of $M_0 > 0$ and $S_0 < 0$, there does exist in $\mathcal{H}$ a unique minimum energy state f^-. f^- depends only on $|x|$ and $|v|$, and, for $p > 9/7$, it is at compact support in R^6.
- f^- is a solution of the Euler-Lagrange system of equations associated with the extremization problem of E in $\mathcal{H}$, i.e.

$$C'[f^-] = K\left(\epsilon_0 - \frac{|v|^2}{2} - \Phi^-\right) \qquad \text{where} \quad f^- > 0 \quad \text{in} \quad R^6\,,$$
$$\Delta\Phi^- = 4\pi G \int f^-\,\mathrm{d}v\,, \tag{9}$$
$$\int f^-\,\mathrm{d}\xi = M_0 \qquad \text{and} \qquad -\int C(f^-)\,\mathrm{d}\xi = S_0\,,$$

  where K and ϵ_0 are unknown constants. In fact, this system turns out to have f^- as its only solution, and then f^- is also the *unique energy extremizer* in $\mathcal{H}$.
- Consider the "conjugate" problem which consists to find a global maximum entropy state at given mass and energy ($M[f] = M_0$ and $E[f] = E_0 < 0$). Then this problem has a unique solution, which is just the f^- which has the right energy, i.e. $E[f^-] = E_0$.
- f^- is "dynamically" linearly stable, i.e. it satisfies the so-called Antonov's criterion (e.g. Binney et al. 1984). It is worth noticing that our arguments here to conclude at the linear stability do not use at all the strong Antonov's constraint on the perturbations: we just need to require that the variations of the distribution function conserve mass and entropy to the first order! Nonlinear stability as defined by Wiechen et al. (1988) is also easily shown to hold for f^-, but Lyapunov's stability (as defined e.g. in Kandrup 1990), although likely, is much more difficult to obtain and we have not yet a complete proof.

5. Conclusion

We have established that a self-gravitating system having prescribed values of its mass and of one of its quasi-entropy satisfying condition (4), admits a unique minimum energy state if $p \geq 9/7$, but not such a state if $p < 9/7$. We have also derived some important properties of the energy minimizers in the case $p \geq 9/7$.

References

Aly J.J., 1989, MNRAS 241, 15
Binney J., Tremaine S., 1984, Galactic Dynamics, Princeton Univ. Press
Ipser J.R., 1974, Astrophys. J. 193, 463
Ipser J.R., Horwitz G., 1979, Astrophys. J. 232, 863
Kandrup H.E., 1990, Astrophys. J. 351, 104
Tremaine S., Hénon M., Lynden-Bell D., 1986, MNRAS 219, 285
Wiechen H., Ziegler H.J., Schindler K., 1988, MNRAS 232, 623

Effective Collision Term Induced by Coarse-Graining

Toshio TSUCHIYA

Department of Physics, Kyoto University, Kyoto 606, Japan

1. Introduction

Since stellar systems such as elliptical galaxies contain a large number of stars, observable quantities are always macroscopic. The macroscopic quantities are obtained by *coarse-graining*. The coarse-graining corresponds to observations and the procedure does not change the evolution of the systems. However, it is important to study apparent effects of coarse-graining, because we always employ coarse-graining in observations and/or N-body calculations.

Lynden-Bell's idea (1967) about the relaxation of collisionless systems is that the coarse-grained system can relax although the microscopic system cannot have increasing entropy. The coarse-graining must play a role for increasing entropy, but it seems doubtful that it is effective for relaxing the system, because the evolution of the coarse-grained system is only an apparent one. In fact, numerical simulations have shown that some equilibriums do not have Lynden-Bell's distribution (van Albada 1982; Tanekusa 1987). Other authors have shown that coarse-grained evolution does not always increase the Boltzmann entropy (Tremaine, Hénon & Lynden-Bell 1986; Kandrup 1987; Mathur 1988; Soker 1990).

In this way, in order to get more basic understanding of the evolution, it is necessary to investigate the characteristics of the collision term induced by coarse-graining.

2. Coarse-Graining

Let us consider a system with N identical point particles which interact with each other only through the gravitational force. A full description of the state of the system is given by specifying the number of particles $Nf(\mathbf{x},\mathbf{v},t)\,\mathrm{d}^3\mathbf{x}\,\mathrm{d}^3\mathbf{v}$ having positions in the small volume $\mathrm{d}^3\mathbf{x}$ centered on $\mathbf{x}$ and velocities in the small range $\mathrm{d}^3\mathbf{v}$ centered on $\mathbf{v}$, at any time t. The quantity $f(\mathbf{x},\mathbf{v},t)$ is called the *distribution function* (DF) of the system. A point particle has density proportional to the delta function, so that f is given by

$$f(\mathbf{w},t) = \frac{1}{N}\sum_{i=1}^{N}\delta\big(\mathbf{w}-\mathbf{w}^{(i)}(t)\big)\,. \tag{1}$$

Here we denote the phase space coordinates $\mathbf{w} = (\mathbf{x},\mathbf{v})$ for brevity, and $\mathbf{w}^{(i)}(t) = \big(\mathbf{x}^{(i)}(t),\mathbf{v}^{(i)}(t)\big)$ are the position and the velocity of the i^{th} particle at time t. Time evolution of the system is given by the following equation

$$\frac{\mathrm{d}f}{\mathrm{d}t} \equiv \frac{\partial f}{\partial t} + \dot{\mathbf{x}}\frac{\partial f}{\partial \mathbf{x}} + \dot{\mathbf{v}}\frac{\partial f}{\partial \mathbf{v}} = \frac{\partial f}{\partial t} + \mathbf{v}\frac{\partial f}{\partial \mathbf{x}} - \frac{\partial \Phi(\mathbf{x})}{\partial \mathbf{x}}\frac{\partial f}{\partial \mathbf{v}} = 0\,. \tag{2}$$

Coarse-graining is defined formally by the expression

$$\langle f\rangle(\mathbf{w},t) \equiv \int \mathrm{d}^3\mathbf{w}_1\, D(\mathbf{w}_1) f(\mathbf{w}+\mathbf{w}_1,t) = \frac{1}{N}\sum_{i=1}^{N} D\big(\mathbf{w}-\mathbf{w}^{(i)}(t)\big)\,. \tag{3}$$

The coarse-grained system can be regarded as a system with N identical smoothed particles with intrinsic density $D(\mathbf{w})$. Note that the particles are smoothed in velocity space as well as coordinate space.

3. The Collision Term Caused by Coarse-Graining

The collision term due to coarse-graining (we refer to it as Γ_{CG}) is defined as the Lagrange derivative of the coarse-grained DF $\langle f\rangle$:

$$\Gamma_{\mathrm{CG}} \equiv \left\langle \frac{\mathrm{d}}{\mathrm{d}t} \right\rangle \langle f\rangle \equiv \left\{ \frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{x}} - \frac{\partial\langle\Phi\rangle}{\partial \mathbf{x}}(\mathbf{x})\frac{\partial}{\partial \mathbf{v}} \right\} \langle f\rangle\,, \tag{4}$$

where $\langle \mathrm{d}/\mathrm{d}t\rangle$ means the Lagrange derivative along the coarse-grained flow of particles.

After straightforward transformation, using Eqs. (2) and (3), Γ_{CG} is rewritten in the form $\Gamma_{\mathrm{CG}} = \Gamma_x + \Gamma_{\mathrm{diag}} + \Gamma_{\mathrm{offdiag}}$, where

$$\Gamma_x = \frac{1}{N}\sum_{i=1}^{N}\big(\mathbf{v}-\mathbf{v}^{(i)}\big)\frac{\partial D\big(\mathbf{w}-\mathbf{w}^{(i)}\big)}{\partial \mathbf{x}}\,, \tag{5}$$

$$\Gamma_{\mathrm{diag}} = -\frac{1}{N}\sum_{i=1}^{N}\int \mathrm{d}^6\mathbf{w}_1 \frac{\partial}{\partial \mathbf{x}}\left(\frac{-Gm}{|\mathbf{x}-\mathbf{x}_1|}\right) D\big(\mathbf{w}_1-\mathbf{w}^{(i)}\big)\frac{\partial D\big(\mathbf{w}-\mathbf{w}^{(i)}\big)}{\partial \mathbf{v}}\,, \tag{6}$$

and

$$\Gamma_{\mathrm{offdiag}} = \frac{1}{N}\sum_{i\neq j}\left\{ G\frac{\mathbf{x}^{(j)}-\mathbf{x}^{(i)}}{|\mathbf{x}^{(j)}-\mathbf{x}^{(i)}|^3} - \int \frac{\partial}{\partial \mathbf{x}}\left(\frac{-Gm}{|\mathbf{x}-\mathbf{x}_2|}\right) D\big(\mathbf{w}_2-\mathbf{w}^{(i)}\big)\,\mathrm{d}^6\mathbf{w}_2 \right\} \frac{\partial D\big(\mathbf{w}-\mathbf{w}^{(j)}\big)}{\partial \mathbf{v}}\,. \tag{7}$$

The first term Γ_x means the rate of change of the intrinsic density of each smoothed particle, due to the shear flow inside their spread. This causes the distortion of the figure in phase space, this effect is regarded as *phase-mixing.*

The coefficients of $\partial D(\mathbf{w}-\mathbf{w}^{(i)})/\partial \mathbf{v}$ are the gravitational accelerations of the i^{th} particle produced by the other particles. Thus Γ_{diag} does not represent the change of the intrinsic density of the i^{th} smoothed particle due to the interaction among the other particles, but due to the potential force produced by the particle itself, so it has to be an artificial effect of the procedure.

Γ_{offdiag} means a very effect of the gravitational interaction. The first term of the coefficient in Γ_{offdiag} is the gravitational force at $\mathbf{x}^{(j)}$ produced by the other point particles. On the other hand, the second term is the force at $\mathbf{x}$ produced by the smoothed particles. Since the second term becomes negligible compared to the first one for particles which lie inside the spreading of the considered particle, the first term gives the *scattering* by neighbouring particles. For distant particles, the difference between the two terms is approximately the difference between the force at a point $\mathbf{x}$ inside the spread of the smoothed particle and its centre $\mathbf{x}^{(j)}$. This moderate change corresponds to phase mixing.

4. Collision Term Derived by BBGKY Formalism

Since BBGKY formalism is one of the most popular method to deal with many body systems (see, for example, Binney & Tremaine 1987), we next discuss the collision term derived by it.

The collision term, Γ_{BBGKY}, is determined by two-point correlation function. BBGKY formalism is based on Liouville theorem in N-dimensional phase space (Γ space). Then we give the N-body DF of the form

$$\left\langle f^{(N)}\right\rangle(\mathbf{w}_1,\ldots,\mathbf{w}_N,t_0) = \frac{1}{N!}\sum_{(i_1,\ldots,i_N)} D\left(\mathbf{w}_1-\mathbf{w}^{(i_1)}(t_0)\right)\cdots D\left(\mathbf{w}_N-\mathbf{w}^{(i_N)}(t_0)\right). \quad (8)$$

where the summation is carried out through all permutations of $(i_1,\ldots,i_N) = (1,\ldots,N)$. This gives the same one-body DF (3) at $t = t_0$. Here we should note that Eq. (8) does not mean coarse-graining, but an ensemble of different states of the system. Thus the collision term derived by BBGKY is different from the one induced by coarse-graining, by definition. The comparison between the two terms, however, is very important.

The two-point correlation function obtained from Eq. (6) is

$$\langle g\rangle(\mathbf{w}_1, \mathbf{w}_2, t_0) = -\frac{1}{N^2}\sum_{i=1}^{N} D\big(\mathbf{w}_1 - \mathbf{w}^{(i)}(t_0)\big) D\big(\mathbf{w}_2 - \mathbf{w}^{(i)}(t_0)\big) + \frac{1}{N^3}\sum_{i\neq j} D\big(\mathbf{w}_1 - \mathbf{w}^{(i)}(t_0)\big) D\big(\mathbf{w}_2 - \mathbf{w}^{(j)}(t_0)\big) \,. \tag{9}$$

Then the collision term is given by

$$\begin{aligned} \Gamma_{\text{BBGKY}} &\equiv \int \frac{\partial}{\partial \mathbf{x}_1}\frac{-GNm}{|\mathbf{x}_1 - \mathbf{x}_2|}\frac{\partial}{\partial \mathbf{v}_1} g(\mathbf{w}_1, \mathbf{w}_2, t)\, \mathrm{d}^6\mathbf{w}_2 \\ &= -\frac{1}{N}\sum_{i=1}^{N}\int \frac{\partial}{\partial \mathbf{x}_1}\frac{-Gm}{|\mathbf{x}_1 - \mathbf{x}_2|} D\big(\mathbf{w}_2 - \mathbf{w}^{(i)}\big)\, \mathrm{d}^6\mathbf{w}_2 \frac{\partial}{\partial \mathbf{v}_1} D\big(\mathbf{w}_1 - \mathbf{w}^{(i)}\big) \\ &\quad + \frac{1}{N^2}\sum_{i\neq j}^{N}\int \frac{\partial}{\partial \mathbf{x}_1}\frac{-Gm}{|\mathbf{x}_1 - \mathbf{x}_2|} D\big(\mathbf{w}_2 - \mathbf{w}^{(j)}\big)\, \mathrm{d}^6\mathbf{w}_2 \frac{\partial}{\partial \mathbf{v}_1} D\big(\mathbf{w}_1 - \mathbf{w}^{(i)}\big) \,. \end{aligned} \tag{10}$$

The first term is the same as Γ_{diag}. As mentioned in the previous section, this term contains only self-interacting effect inside the smoothed particles, which cannot express the "collisions" among the particles. The second term shows the interaction among different smoothed particles like the one we saw in Eq. (7), but the term derived here has the factor $1/N^2$. Therefore this interaction term is negligible compared with that induced by coarse-graining.

5. Discussion

We have shown that three different effects come in the collision terms: 1) the self-interaction, 2) the phase-mixing, and 3) the scattering. The self-interaction term is only an artificial effect of the coarse-graining procedure and has no physical meaning. The scattering term is the usual idea about collision term, but the phase-mixing term arises from the long range force characteristics of gravity.

With the same initial distribution as a coarse-grained system, the collision term obtained by using the two-body correlation function contains only the self-interaction term and a weaker interaction term than the coarse-graining one.

In BBGKY formalism, the N-body DF (Eq. (8)) is conserved along the motion. On the other hand, in the coarse-graining procedure we can define the coarse-grained N-body DF which has initially the same form as Eq. (8); however, the evolution is different. As the coarse-grained DF (Eq. (3)) changes in time, the coarse-grained N-body DF also changes. That is the reason why the collision term Γ_{BBGKY} is negligible compared with the one induced by coarse-graining.

The BBGKY formalism is concerned with the *statistical average* of an ensemble of different states of the system. If the system is ergodic, the average

corresponds to a long time average. For systems which evolve thermodynamically, such as globular clusters, we can deal with the time average, thus the BBGKY formalism is appropriate. For elliptical galaxies, however, the evolution that we usually consider is dynamical, thus we cannot define any time average. With the result, in studying the dynamics of elliptical galaxies, that the collision term should be estimated by the one directly derived by coarse-graining.

References

Binney J., Tremaine S., 1987, Galactic Dynamics, Princeton, Chap. 8
Kandrup H.E., 1987, MNRAS 225, 995
Lynden-Bell D., 1967, MNRAS 136, 101
Mathur S.D., 1988, MNRAS 231, 367
Soker N., 1990, Publ. Astron. Soc. Pac. 102, 639
Tanekusa, 1987, J. Publ. Astron. Soc. Japan 39, 425
Tremaine S., Hénon M., Lynden-Bell D., 1986, MNRAS 219, 285
Tsuchiya T., 1993, Prog. Theor. Phys., to be published
van Albada T.S., 1982, MNRAS 201, 939

Theoretical and Numerical Investigation of the Stability of Flattened Galaxies

E. GRIV & W. PETER

Department of Physics, Ben-Gurion University, 84105 Beersheva, Israel

Abstract. Quasi-linear theory is applied to the wave-star interaction of a differentially-rotating stellar disk of a galaxy. Under the influence of growing spiral waves the velocity dispersion of stars increases, and the resulting distortion in phase space leads to a decrease in the growth rate of the waves, and the Jeans instability ends. Due to interactions of stars with unstable waves the relaxation of the stellar disk of the Galaxy occurs in $\sim 10^9$ years. The theory is confirmed by N-body computer simulations.

In modern density wave theory, the spiral structure of galaxies results from a non-axisymmetric gravitational Jeans instability (see Lin & Bertin 1984 for review). Up to now, density wave theory has been developed only in the linear approximation. Weak nonlinear theory, i.e., nonlinear theory using a perturbation approach, can be used to explain a broad class of phenomena in the stellar disk of a galaxy. This approach was initially developed for the physics of an inhomogeneous plasma in a magnetic field (Krall & Trivelpiece 1973). For instance, under the action of growing waves, the average properties of a stellar disk may change; the "temperature" of the system which is measured by the kinetic energy of random motion may increase. This increase in turn leads to a stabilization of the gravitational instability as computer experiments have shown (Hohl 1972; Miller 1976).

To describe the properties of stellar disks in the nonlinear regime it is useful to consider the *quasi-linear wave-star interaction* that is not associated with the resonance condition $\omega = kv$, where ω is the wave frequency, k is the wavenumber, and v is the velocity.

Let us consider an inhomogeneous differentially-rotating stellar disk of a galaxy, taking a kinetic description as a basis. To describe the problem, it is useful to employ the well-developed mathematical formalisms in plasma physics (e.g., quasi-linear theory) which have dealt with similar problems. For a discussion on the formal analogy between the oscillations of a differentially-rotating stellar disk and the oscillations of a magnetic plasma see the mono-

graph by Fridman & Polyachenko (1984). By using the concepts of quasi-linear theory we can show that through the influence of growing spiral waves the stars will tend to diffuse in velocity and coordinate space according to the relation:

$$\frac{\partial f_0}{\partial t} = \alpha \frac{\partial^2 f_0}{\partial \mathbf{v}^2} - \beta \frac{\partial^2 f_0}{\partial \mathbf{r}^2} \tag{1}$$

Here we have presented the distribution of stars in the form $f(t) = f_0(t) + f_1(t)$, where $f_0(t)$ changes slowly in time and the small perturbation $f_1(t)$ changes rapidly. See Krall & Trivelpiece (1973), Griv (1992), and Grivnev (1988) for a more detailed derivation of Eq. (1). This equation has the following solutions for the non-resonant (or adiabatic) interaction of waves and stars:

$$f_0(v) \sim \frac{1}{\sqrt{\sigma_0^2 + |\gamma|\mathcal{E}}} \exp\left[\frac{-v^2}{2(\sigma_0^2 + |\gamma|\mathcal{E})}\right] \tag{2}$$

and

$$n(r) \sim \frac{1}{\sqrt{n_0^2 - |\xi|\mathcal{E}}} \exp\left[\frac{-r^2}{2(n_0^2 - |\xi|\mathcal{E})}\right] \tag{3}$$

In these equations, $n(r)$ is the surface mass density, σ_0 (n_0) is the initial velocity (surface density) dispersion, and $\mathcal{E}$ is the square of the wave amplitude.

According to Eqs. (2-3), the main body of the distribution is effectively heated and the surface density is redistributed by the unstable waves. The velocity dispersion of the non-resonant part of the distribution function increases and $f_0(v)$ becomes *less peaked*, and the surface density becomes *more peaked* as the wave energy increases. The diffusion of the stars takes place because the stars gain additional oscillatory energy in the gravitational field of the unstable density waves.

According to Eqs. (2-3), under the action of growing waves the velocity dispersion of stars will increase and the density of the disk falls off exponentially. This distortion in phase space caused by the spiral waves leads to a decrease in the growth rate of the waves and eventually to the cessation of the instability. Therefore, from the theoretical point of view, the spiral waves in computer-generated galaxies have to be short-lived, and should dissipate after a few rotations of the disk system.

Our theoretical results have been verified by three-dimensional N-body simulations. We investigated the evolution of a model for an isolated thin disk of a galaxy by direct integration over a time span of equation of motion of identical stars. The cutoff radius r_c of the Newtonian potential was introduced in order to eliminate close encounters between the model particles. This "softening" parameter reduces the interaction at short ranges and puts a lower limit on the "size" of the particles.

At the start of the N-body integration a mass density variation given by

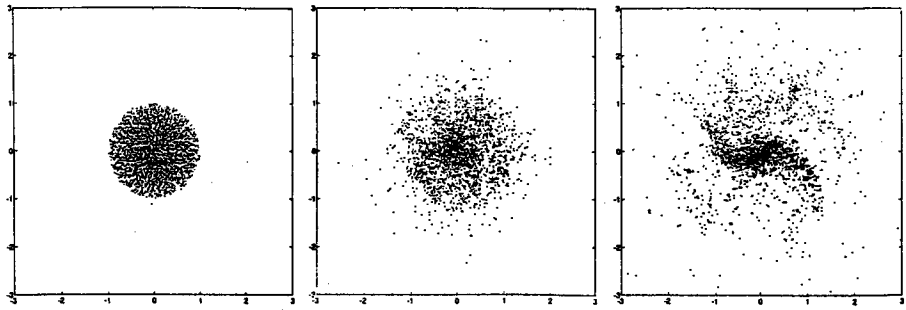


Fig. 1. The evolution of the thin disk of a galaxy with $N = 3186$ stars. Snapshots are taken at normalized times (left) $t = 0$, (middle) $t = 400$, and (right) $t = 1000$. The time is normalized so that the time $t = 1000$ corresponds to a single revolution of the initial disk. The central dense core and under dense "corona" is due to the gravitational Jeans instability. Note the large velocity dispersions of the particles at the end of the calculation

$$n(r) = n(0)\left(1 - \frac{r^2}{R^2}\right)^{1/2} \tag{4}$$

was used, where $n(0)$ is the central surface density and R is the radius of the disk. To ensure initial equilibrium of the disk the uniform angular velocity

$$\Omega_0 = 0.808\pi \left(Gn(0)/2R\right)^{1/2} \tag{5}$$

was adopted (Hohl 1972). Then the position of each particle was slightly perturbed by applying a pseudo-random number generator. The Maxwellian-distributed random velocities with radial and azimuthal dispersions according to the well-known Toomre criterion

$$\sigma_0 = 0.341\Omega_0 \left(R^2 - r^2\right)^{1/2} \tag{6}$$

were added to the initial circular velocities (Hohl 1972). So initially the disk is Jeans stable against the small-scale axisymmetric perturbations but unstable against the slow-growing large-scale spiral perturbations (see Peter et al. 1993, for a more detailed explanation).

Corrections are then applied to the resultant velocities and coordinates of the model stars so as to ensure the equilibrium between the centrifugal and gravitational forces and to preserve the position of the disk center of gravity at the origin.

The sense of disk rotation was taken to be counterclockwise and units are such that $G = 1$, $R = 1$, $r_c = 0.0025$ and the mass of each star $m = 1$. A time $t = 1000$ is taken to correspond to a single revolution of the initial disk.

In Fig. 1 we show a series of three snapshots from a three-dimensional simulation ($N = 3186$) run on a Cray YMP supercomputer. It is seen that the effects of the gravitational Jeans large-scale instability appear quickly in the

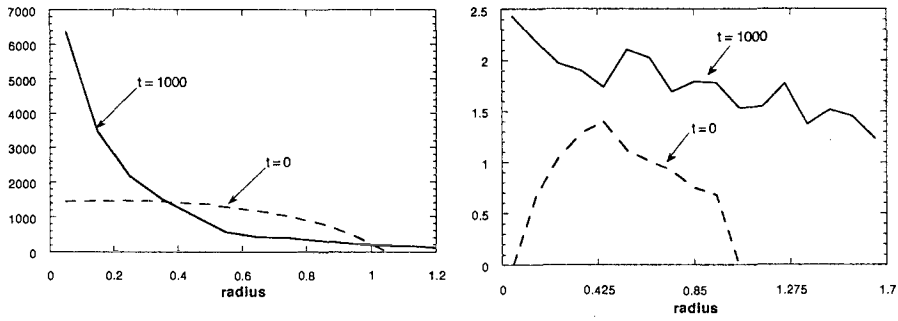


Fig. 2. N-body simulation ($N = 3186$) results for (left) surface mass density as a function of radius, and (right) radial dispersion velocity vs. radius at times of $t = 0$ and $t = 1000$. All units are arbitrary. The results are in agreement with the theory presented here

simulation. At first, a tightly-wound spiral structure appears, but this gives way to a more open spiral structure with a larger pitch angle. In accordance with the given above theoretical conclusion the unstable spiral waves (the spiral arms) are short-lived and exist only during the time of $2 - 3$ rotations of the system. As is seen in Fig. 2, the velocity dispersion increases and an exponential disk develops under the action of the gravitational instability, in accordance with the theory presented here.

Acknowledgements. E.G. has been supported by the Israeli Ministry of Absorption and W.P. has been supported by the Israeli Ministry of Science. Supercomputer calculations were supported by a grant from the San Diego Supercomputer Center which is funded by the National Science Foundation.

References

Fridman A.M., Polyachenko V.L., 1984, Physics of Gravitating Systems, Springer, New York

Griv E., 1992, in ESO/EIPC Workshop on Structure, Dynamics and Chemical Evolution of Early-Type Galaxies, eds. I.J. Danziger et al., ESO, Garching, p. 207

Grivnev E.M., 1988, Sov. Astron. 32, 139

Hohl F., 1972, J. Comput. Phys. 9, 10

Lin C.C., Bertin G., 1984, Advances in Appl. Mechanics 24, 155

Krall N.A., Trivelpiece N., 1973, Principles of Plasma Physics, McGraw-Hill, New York

Miller R.H., 1976, J. Comput. Phys. 21, 400

Peter W., Griv E., Peratt A.L., 1993, in Progress in New Cosmologies, eds. H. Arp, C.R. Keys, K. Rudnicki, Plenum Press, Cambridge, (to be published)

The Evolution of Orbits in the Stellar Disk as a Purely Discontinuous Random Process

I.V. PETROVSKAYA

St Petersburg State University, Bibliotechnaya pl. 2, St Petersburg, 198904, Russia

Abstract. The scheme of a purely discontinuous random process was applied to the investigation of irregular forces effects (Petrovskaya 1969a). Earlier, Hénon (1960) showed that there are some features which could be described only by such a scheme. Later, the scheme was applied for studying the evolution of the velocity distribution of star groups of different masses in the uniform 3-D stellar field (Petrovskaya, 1969b; Kaliberda & Petrovskaya 1970, 1971, 1972).

In the purely discontinuous scheme the characteristics of motion change as jumps and that gives us the possibility to evaluate not only the mass, but also the energy taken away by dissipated stars. In this study we consider the evolution of orbit integrals under the action of stellar-stellar encounters. So we have to take into account the smoothed potential of the system. We use the Kolmogorov-Feller's equation describing the time evolution of the distribution function of two orbit integrals, the energy integral, H, and the angular momentum integral, Q. The kernel of the equation depends on the function $F(h, q, H, Q)$, which is equal to the density of the probability that as a result of the encounter the test star and the field star the value of energy integral of the test star changes from h to H and the the value of integral momentum changes from q to Q. We obtained here the expression for $F(h, q, H, Q)$ for the case when the system is flat and the mass of the test star is zero. This probability function was multiplied by the coefficient of multiplicity for the flat medium found earlier by Petrovskaya, Chumak, & Chumak (1984).

So the result probability is found to be finite everywhere. The latter gave us the opportunity to apply the method of integral transformation for solving the Kolmogorov-Feller equation, as it was suggested earlier (Petrovskaya 1983).

References

Hénon M., 1960, Ann.d'Ap. 23, 459
Kaliberda V.S., Petrovskaya I.V., 1970, Astrofiz. 6, 135
Kaliberda V.S., Petrovskaya I.V., 1971, Astrofiz. 7, 663
Kaliberda V.S., Petrovskaya I.V., 1972, Astrofiz. 8, 305
Petrovskaya I.V., 1969a, Astron.Zh. 46, 824
Petrovskaya I.V., 1969b, Astron.Zh. 46, 1220
Petrovskaya I.V., 1983, Vestn. Leningradsk. Univ. No. 13, 94
Petrovskaya I.V., Chumak Z.N., Chumak O.V., 1984, Astron.Zh. 61, 467

7. Galactic and Extragalactic Problems

Interacting Spherical Stellar Systems

Juan C. MUZZIO [1 2]

[1] Facultad de Ciencias Astronómicas y Geofísicas - UNLP
[2] PROFOEG - CONICET

Abstract. Spherical stellar systems offer what is probably the best example of regular stellar systems and, accordingly, quite a few models are available for them with distribution functions depending on energy alone or on both energy and angular momentum. Real stellar systems, however, are subject to several effects that, although ignored when deriving those models, can cause irregular orbits to appear in them. Among those effects, the interaction of stellar systems is an extreme example of phenomena that, except for a few general analytical results, can only be investigated through numerical methods. While recognizedly a difficult subject, the study of interacting spherical stellar systems offers an interesting and almost virgin field to chaos researchers. We summarize here the investigations of other authors and our own, with a detailed discussion of the different physical processes relevant to this problem, and with particular emphasis on the most recent results.

1. Ideal Versus Real Spherical Stellar Systems

A star in a static spherical potential moves on a plane and the typical bound orbit forms a rosette (see, e.g., Landau & Lifshitz 1960, pp. 30–35). Since two integrals of motion (namely, energy, E, and angular momentum, L) are available, the motion is regular. Therefore, the distribution function (DF) of collisionless spherical stellar systems depends in general on the energy and the magnitude of the angular momentum, $f(E, L)$; when the DF depends on energy alone, the velocity dispersion tensor for the system is isotropic (Binney & Tremaine 1987, pp. 221–223). Spherical stellar systems with isotropic velocity distributions are stable under fairly general conditions, while the stability of systems with anisotropic velocity distribution is not so well understood (see, e.g., Binney & Tremaine 1987, pp. 302 to 308); spherical systems with almost radial orbits, in particular, are subject to the so called *radial orbit instability* that turns them into triaxial systems (see, e.g., Merritt 1987).

While *ideal* collisionless spherical stellar systems may thus offer the best possible example of regular stellar systems, *real* systems are subject, instead,

to different degrees of chaotic behavior. Núñez et al. (1993) showed that, provided the orbital angular momentum is low enough, even very weak bar-like perturbations can give rise to irregular orbits in spherical potentials, and the oscillations found by Miller (1993) offer another path to chaotic behavior in spherical stellar systems. Irregular orbits can also easily appear in interacting stellar systems. They were found, for example, in the restricted three-body problem by Hénon (1966) and Jefferys (1966). An example closer to galactic dynamics can be found in the model of Rix & White (1989), although in that case the galaxies were ellipsoidal (rather than spherical) and immersed in a common envelope. Besides, close encounters and mergers of spherical galaxies often result in triaxial, rotating, systems where chaotic orbits must be present.

Therefore, we may conclude that *real* spherical stellar systems are subject to perturbations that easily turn them into *irregular* systems, and that the interaction of spherical galaxies offers a "path to chaos" that has not been much explored thus far. Our purpose here is to present an interesting, almost virgin, field of research to colleagues working on chaos and ergodicity. A caveat should be made that this is not an easy subject, and its difficulties may help to understand why so little has been done on the study of chaos in interacting stellar systems. Alternatively, the presence of irregular orbits in stellar systems is not a mere oddity, but a matter of considerable interest to understand what happens to the system, as underlined in particular by the work of Pfenniger (1986). He showed that such effects as relaxation and dynamical friction are greatly enhanced by the presence of irregular orbits in the stellar system, and he also indicated interacting galaxies as a possible example of systems where these phenomena could be of relevance (see, also, Udry & Pfenniger 1988).

The earlier work on interacting galaxies was reviewed by White (1982), and more recent reviews are those of Muzzio (1987) and Barnes & Hernquist (1992). Consequently with our purpose, in the present review we will only consider the simplest possible theoretical case, that is, collisionless, non rotating, spherical stellar systems (no gas, no disks, no observational comparisons; see Barnes & Hernquist 1992 for those subjects). We will pay special attention to the physical processes relevant to the study of interacting stellar systems and, although references to earlier work are unavoidable, we will concentrate on the most recently published results and to work in progress, plus a few papers that seem to have escaped the attention of previous reviewers. The case of satellites (that is, one of the systems much smaller than the other) is particularly interesting to show the influence of the different physical processes at work, and will be also discussed at some length.

2. Dynamical Processes

Several dynamical processes play a significant role in the interaction of stellar systems and, since they may be of particular interest to researchers from fields other than stellar dynamics, they will be discussed at some length in this section.

2.1 Relaxation

An excellent example of the importance of relaxation effects in the interaction of stellar systems is provided by Binney & Tremaine (1987, p. 435): a stellar system, initially in equilibrium, is perturbed by a fast encounter with another system, suffers a sudden increase of its kinetic energy and, after a relaxation period, returns to an equilibrium state different from the original one; the change of kinetic energy during the relaxation phase, however, turns out to be *twice* the original change caused by the encounter.

Nevertheless, relaxation is not *exclusively* present in the case of *interacting* stellar systems and it is thoroughly discussed by other authors in the present volume, so that we will not discuss it any further. Suffice here to note that some of the results we present below are based on the classical two-body estimates, so that relaxation effects there may have been grossly underestimated as compared with the "collective" relaxation effects of Gurzadyan & Savvidy (1986), or with the response of irregular orbits (Pfenniger 1986; Udry & Pfenniger 1988).

2.2 Dynamical Friction

Chandrasekhar (1943) considered particle-particle interactions to derive the equation for the deceleration suffered by a particle of mass M that moves, with velocity V, through an infinite homogeneous medium of particles of mass m. If the density of the medium is ρ, its velocity distribution is Maxwellian with dispersion σ and G is the constant of gravitation, his result for the deceleration is:

$$\frac{\mathrm{d}V}{\mathrm{d}t} = -4\pi G^2 \frac{(M+m)\rho \ln \lambda}{V^2} \left[\mathrm{erf}(X) - \frac{2}{\sqrt{\pi}} X \exp(-X^2) \right] \tag{1}$$

where $X = V/(\sigma\sqrt{2})$ and $\ln \Lambda$ is the "Coulomb logarithm":

$$\Lambda = \frac{b_{\max} V^2}{G(M+m)} \tag{2}$$

that needs a *maximum* distance cutoff, $b_{\max}$ (the inclusion of a *minimum* distance cutoff in some derivations of Chandrasekhar's equation is just an artifact, as shown by Muzzio & Vergne 1988). The proper election of the $b_{\max}$

value has long been the object of debate, the most recent contribution to the subject being the one by Smith (1992). The truth is that only *close* encounters are "instantaneous", as Chandrasekhar regarded them in his derivation; Hénon (1958, 1960) recognized the problem and improved the formula with a perturbation treatment of distant encounters. His results are larger than those of Chandrasekhar by about 50%.

Following the approach of Kandrup (1983), Bekenstein & Maoz (1992) considered the action of the stochastic force of the background on the test particle, rather than a particle-particle approach. Their formula agrees with the one of Chandrasekhar for massive test-particles, but extends it for other background- to test-particle mass ratios, $\beta = m/M$:

$$\frac{\mathrm{d}V}{\mathrm{d}t} = -4\pi G^2 \frac{M\rho \ln\lambda}{V^2} \left[\mathrm{erf}(X) - \frac{2}{\sqrt{\pi}}(1+\beta)\, X \exp(-X^2)\right] \tag{3}$$

They also obtained the velocity dispersion induced on the test-particle (which Hénon had found too). Some curious properties of dynamical friction were recently investigated by Zamir (1992).

Kalnajs (1972) rederived Chandrasekhar's formula considering the *drag force* exerted on the test-particle by its "wake". This approach leads to the (apparently) absurd result that the first- (or even second-) order effect of dynamical friction on a test-particle moving through a disk (Kalnajs 1972) or a sphere (Tremaine 1981) is zero. The key to solve that paradox lies in the fact that *resonances* are crucial to dynamical friction in those cases (Tremaine 1981; Tremaine & Weinberg 1984). Detailed analyses of the orbital decay of a satellite within a spherical galaxy are due to Weinberg (1986, 1989).

A very interesting question is whether dynamical friction can produce *dynamical heating*. The orbital energy of a globular cluster is about 1,000 times its internal (i.e., binding) energy: if there were a process (even a very inefficient one) capable of introducing even a very small part of the braking energy inside the cluster, that would play havoc in its structure! This very important point was raised by Miller & Smith (1985) and the big question is, of course, whether such a process exists.

Muzzio et al. (1988) used a *linear* analysis to show that the tidal force exerted by the wake on the satellite is negligibly small. Therefore, if the differential effect of the tidal force produces the dynamical heating, the latter cannot be significant; alternatively, if dynamical heating actually exists, it may have an origin different from the tidal force of the wake. Without considering the possible origin of dynamical heating, Muzzio & Plastino (1992) used a purely experimental approach and found from numerical N-body simulations that their satellite could absorb no more than 5% of the braking energy (and, probably, less than that).

Nevertheless, since even a 0.1% efficiency can be relevant, more detailed investigations of the subject are warranted. The investigation of Muzzio et al. (1988) has to be improved using non-linear analysis, and numerical simu-

lations with larger numbers of bodies would be useful to derive tighter upper limits for possible dynamical heating effects.

2.3 Tidal Effects

According to common wisdom, a satellite orbiting a larger body gets its material lying outside the Roche lobe pruned. King's (1962) idea of imposing a *tidal radius*, r_t on a globular cluster, or satellite galaxy, was a very important and useful one. For a cluster of mass M, orbiting a galaxy of mass M_g with perigalactic distance R_p, the tidal radius is:

$$r_t = R_p \left(\frac{M}{3.5 M_g} \right)^{1/3} \tag{4}$$

Nevertheless, tidal radii are much less well defined than the Roche lobe idea suggests. For example, stars on retrograde orbits around the satellite can remain stable at much larger distances than the tidal radius (Innanen 1979), because the Coriolis force, which is always directed along the satellite's radius, has inward sense (adding to the satellite's gravity) for retrograde motion, but outward sense (tending to disrupt the satellite) for prograde orbits. Thus, the satellite is much more easily stripped of stars on direct (prograde) orbits than on retrograde orbits. Besides, Grillmair (1992) has noted the presence of a halo of extra-tidal stars which can significantly alter the appearance of the profile and complicates the measurement of tidal radii.

Except for the limiting effect of the tidal radius, the action of the external tidal field on the stars is usually not taken into account for models of globular clusters and similar stellar systems but, very recently, Heggie & Ramamani (1993) generalized King's models using the Jacobi's integral instead of the energy.

Another important effect is *tidal torquing*, discussed by McGlynn & Borne (1991), which contrary to common wisdom does not depend on the satellite being deformed. Let us consider a satellite orbiting around a galaxy, and a star moving around the satellite's center in the same plane of the orbit. Due to symmetry, on first approximation, the effect of direct and retrograde torques cancels along a complete revolution around the cluster. On second approximation, however, as the satellite moves around the galaxy while the star moves around its center, we notice that the motion of the star *relative to the torque* is much faster for a star on a retrograde orbit than for one on a prograde orbit, so that the torquing effect is larger on the latter. Thus, the torque accelerates the prograde star, and decelerates the retrograde one, producing a net (second order) acceleration on the prograde direction. In a way, the effect may be likened to the wave-particle interaction that produces Landau damping (Saslaw 1985, pp. 105–107).

2.4 Effects of Encounters

Let us consider a stellar system of mass m_c and mean square radius r_c that suffers an encounter with a second one of mass m_n. If the relative motion of the two systems is fast enough, the effect of the encounter on the structure of the first one can be investigated using the *impulse approximation* of Spitzer (1958) which assumes that the stars that make up the system do not move during the encounter. Thus, the potential energy of the system remains unchanged while each one of its stars gets a sudden impulse that alters its velocity and, therefore, the total kinetic energy by an amount:

$$\Delta T = \frac{1}{3} m_c \left(\frac{2Gm_n}{p^2 V} \right)^2 r_c^2 \tag{5}$$

where V is the relative velocity of the two systems and p the impact parameter. Some stars may acquire enough energy to escape from the system (a process dubbed *tidal stripping*), and the rest of the stars subsequently relax to a new equilibrium state where the potential energy is equal to twice the kinetic energy.

For slow encounters the *adiabatic approximation* (also pioneered by Spitzer) rather indicates that the stellar orbits in one system react as a whole to the perturbation caused by the other system. Even encounters that are fast compared with the orbital motion of the outermost stars may be slow compared with that of the innermost stars, and Fig. 1 of Aguilar & White (1986) beautifully shows how the outskirts of a stellar system follow the scenario of the impulse approximation, while its nuclear region reacts almost as a solid body according to the adiabatic prescription.

What happens to the material tidally stripped from the galaxies that suffered an encounter? Most of it escapes from the system but, could at least some of the stars that escape from one of the galaxies be captured by the other one? That is, could we have *tidal accretion* besides tidal stripping? The answer is yes. The effect was noted, e.g., by Dekel et al. (1980), but it was regarded by them as being of little relevance. Forte et al. (1982) and Muzzio et al. (1984) showed, instead, that it can be a significant process in the evolution of systems of globular clusters in clusters of galaxies (hence, the name "*cluster swapping*"; see, Muzzio 1988), and Muzzio (1986) showed the same for other galactic material (hence, "*tidal accretion*"). The effect is maximized for slow encounters, because the velocity of the galaxy that accretes the tidally stripped material cannot be too different from the velocity of that material. Up to 8% of the mass of a galaxy can be accreted (Vergne 1992) and rates of 3% – 5% are common (Carpintero et al. 1989). The total effect is of course much larger for successive encounters (about six per Hubble time for a typical galaxy in a cluster, according to Muzzio 1987).

For very slow encounters, the (negative) internal energy of the galaxies can absorb the (negative, zero, or slightly positive) orbital energy of the relative motion (Alladin 1965). The bulk of the material of the two galaxies thus

merges into a single remnant, while the rest of it escapes, helping the system to get rid of any remaining energy excess. Beautiful examples of mergers are shown, e.g., by Villumsen (1982) in his Figs. 2 through 5.

3. Methods

Both analytical and numerical methods are used to investigate the interaction of stellar systems. The former allow the derivation of general results, but are usually limited to the study of highly idealized models, while the latter can be used to investigate conditions closer to reality at the price of considering them on a case by case basis.

3.1 Analytical Methods

In addition to demanding strong simplifying assumptions, analytical methods are usually much more difficult to apply than computer simulations. Despite the necessary approximations, results can be surprisingly accurate, however. The impulse approximation, for example, was found by Aguilar & White (1985) to give very good results even for cases beyond its expected range of application. Excellent agreement between analytical and numerical results was also found by Cincotta et al. (1991) for the probability of escape from a stellar system due to a fast head-on encounter.

Some brave attempts, like those by Knobloch (1978), Weinberg (1986) and Sridhar & Nityananda (1990) can be mentioned, but the field is clearly dominated by computer simulations.

3.2 Numerical Experiments

Most straightforward (and CPU time consuming!) are the *direct summation* methods, where the total force on every particle is computed adding all the forces exerted on that particle by each one of the other particles (see, e.g., Aarseth 1985). Some speed gains may be obtained when part of the particles can be taken as massless (see, e.g., Muzzio et al. 1984).

Tree codes (e.g, Barnes & Hut 1986) use tree-structured data and approximate the force from distant particles with low-order multipole expansions, demanding shorter computation times than the direct summation methods for large numbers of bodies.

The *expansion of the potential on different basis functions* is very popular for investigations of interacting systems (see, e.g., Villumsen 1982; Hernquist & Ostriker 1992). As long as the particle-particle interactions are not relevant to the problem in question, and as long as the galactic potentials can be approximated with a reasonably small number of terms in the expansion, it offers a fast and efficient method.

Methods that distribute the particles in a fixed grid and use FFT techniques to compute the potential (see, e.g., Miller 1978), instead, are less frequently used to investigate collisions of galaxies, because much of the grid is then left empty. The moving grid of Miller (1986) may be useful to circumvent that difficulty.

When studying a small perturbation to a stellar system, only a fraction of the particles is used for the perturbation itself. In such cases, it is much more efficient to represent the system with analytic functions and to use "perturbation particles" of variable mass to simulate the perturbation, as done by Leeuwin et al. (1993) and Wachlin et al. (1993).

Some researchers are now building their own special purpose computers for N-body simulations, like the "GRAvity PipEs" (GRAPEs) at the University of Tokio (Ito et al. 1991; Fukushige et al. 1991; Makino 1991) that allow the use of large numbers of bodies in investigations of galaxy interactions (see, e.g., Okumura et al. 1991).

4. Results

The results obtained by different authors, mainly from numerical simulations, are reviewed in the present section.

4.1 Encounters of Stellar Systems

Dekel et al. (1980) investigated the structure of systems that had suffered slow hyperbolic encounters. As a result, their systems developed, or extended, an inner flat mass profile while the outskirts were stripped. With some exceptions, the innermost regions tended to concentrate while the outer regions expanded. Successive encounters stripped off stars puffed up into the halo by a previous encounter.

Aguilar & White (1985, 1986) found good agreement between the results of the impulse approximation and those of their numerical experiments, and that de Vaucouleurs density profiles were robust (even after 40% mass losses), while King profiles evolved toward de Vaucouleurs ones after the encounter. They also found that tidal effects can only be recognized shortly after an encounter (Fig. 2 of their 1986 paper is particularly interesting).

Vergne (1992) reported mass losses between 3% and 18%, with accretions between 1% and 8%, for equal mass encounters; mass losses can reach 36%, again with about half of that amount being accreted, for $M_2/M_1 = 1/4$. She found some flattening toward the orbital plane for the low mass members.

4.2 Mergers

Vergne (1992) found, for various mass ratios, the regions of the energy vs. impact parameter plane where mergers do and do not occur. In addition to allowing (as previous authors did) infinite time for the merger to take place, she also considered that the Hubble time posed a limit to possible mergers.

The main conclusions derived from the earlier work (already reviewed by Muzzio 1987 and Barnes & Hernquist 1992) are:

a) Mass losses are moderate, so that acceptable estimates of mean radii (say, half-mass radii) and velocity dispersions may be obtained assuming mass and energy constancy, and using the virial theorem;
b) The homology assumption is wrong, however, since merger remnants tend to have higher central densities than their progenitors;
c) The binding energy hierarchy is not destroyed (centers tend to mix with centers, and outskirts with outskirts), so that population gradients are reduced but do not disappear;
d) Remnants tend to be prolate (as a result from head-on encounters), slowly rotating triaxial figures (for small, non-zero, impact parameters), or faster rotating oblate figures (for larger impact parameters), and they are frequently surrounded by an extended envelope where the density falls as r^{-4}.

More recently, Okumura et al. (1991) used their GRAPE-1 system to simulate mergers of systems made up of 16,384 bodies, thus avoiding numerically enhanced relaxation. Their results illustrate beautifully several of the previous conclusions summarized above. Besides, they were able to obtain good rotation curves for their remnants and found that their ratios of maximum rotation velocity to central velocity dispersion were in good agreement with the observational results for elliptical galaxies. They also found good agreement with the Faber-Jackson relation for elliptical galaxies (Faber & Jackson 1976), so that their work gives support to the idea that elliptical galaxies originate from mergers (see, e.g., Toomre 1977).

The need to use large numbers of bodies in merger simulations was also emphasized recently by Gelato et al. (1992).

4.3 Satellites

The orbital decay, due to dynamical friction, of a satellite on a circular (turned into spiral) orbit within a spherical galaxy has been the subject of many investigations. Lin & Tremaine (1983), Bontekoe & van Albada (1987), and Zaritsky & White (1988) (which corrected White 1983) all found good agreement of their numerical experiments with Chandrasekhar's formula and concluded that dynamical friction may be regarded as a *local* processes.

Numerical experiments by Hernquist & Weinberg (1989) resulted in excellent agreement with the decay rates of the other authors, but with a 50%

difference with the predictions of the perturbative analysis of Weinberg (1989) that they attributed to second order effects; they did find good agreement between the numerical and analytical results on the wakes of the satellites, however.

A caveat is needed here: since the deceleration caused by dynamical friction is proportional to the satellite's mass, almost all the studies cited above used satellites one-tenth as massive as the galaxy in order to have a measurable effect, while the real situation (say, globular clusters or dwarf galaxies) calls rather for mass ratios in the $10^{-6} - 10^{-3}$ range. For such low mass ratios, dynamical friction is only relevant in the innermost regions of the galaxy, where the density is high enough, as Fig. 7 of Pesce et al. (1992) dramatically shows. An orbital decay characterized by an essentially constant pericentric distance and a slowly decreasing apocentric distance, as shown in Fig. 1 of Bontekoe & van Albada (1987) is, thus, much more akin to the real situation than the more common view of an almost circular inward spiraling (which can only happen after the very long interval needed for the apocenter distance to become similar to the pericentric one). The fact that real satellites have much lower masses than the values used for the theoretical investigations also casts doubts on the applicability of the global response studies: with satellite-to-galaxy mass ratios of one-tenth those studies make sense, but with mass ratios of one-millionth it is pointless to look for global responses or resonances that would be distorted or erased by other disturbances (density fluctuations within the galaxy, other satellites, and so on). A purely local estimate, like the one of Chandrasekhar, may offer a more sensible approach in those cases.

It is worth recalling, however, that Chandrasekhar's approximation considers the motion in an infinite homogeneous medium. Within a galaxy, instead, the gravitational field of the galaxy itself has to be taken into account too and, for example, tidal accretion by the satellite becomes possible. A new study, closer to the real situation, is now under way at La Plata (Cora et al. 1993).

While the dynamical friction investigations usually take the satellite as a point mass, it is obviously very interesting to know what happens to a real satellite made up of stars and suffering relaxation, tides, etc. Oh & Lin (1992) used a new code developed by Oh et al. (1992) to investigate the evolution due to relaxation and tidal effects (they also included disk shocking, which lies outside the scope of the present review). Their simulations are not self-consistent, however, as the cluster's field is always spherical and constant in time.

The self-consistent simulations of Miller (1986) show dramatically how the satellite evolves along the orbit (see his Fig. 1). Miller found that, after a flyby close to the center of a larger system, the satellite suffers either little mass loss ($< 1\%$) or is totally disrupted. McGlynn (1990), instead, found only negligible effects on the core of a system stripped of as much as 30% of

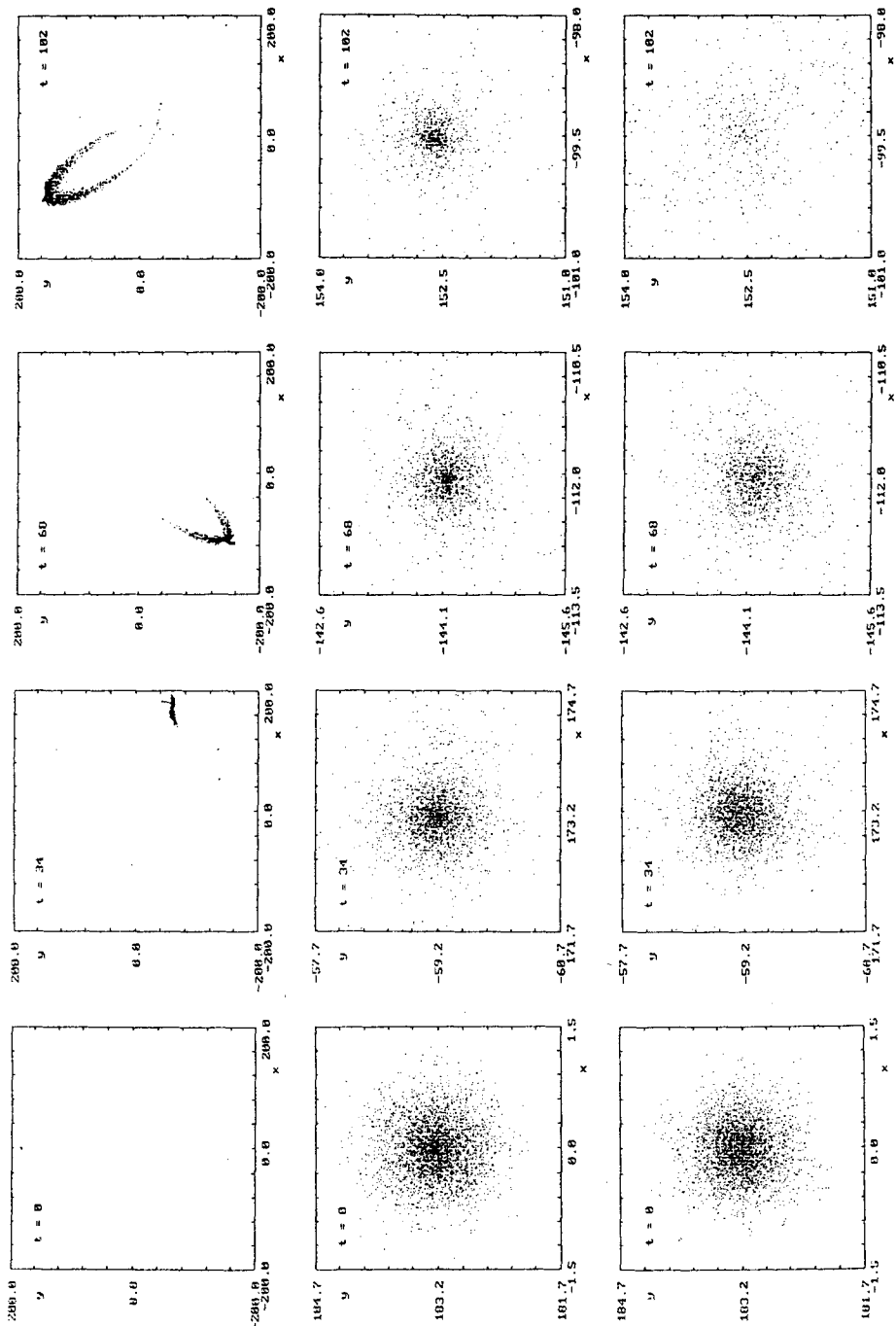


Fig. 1. Disintegration of a dwarf galaxy orbiting a giant one

its outer parts. The cause of the discrepancy is, probably, that McGlynn's King models are much more concentrated than Miller's $n = 3$ polytropes.

Zinnecker et al. (1988) and Freeman (1990) suggested that nucleated dwarf elliptical galaxies, captured by giant elliptical galaxies, might lose their outer envelopes and their nuclei might thus become globular clusters. Their suggestion was criticized by van den Bergh (1990) on observational grounds but, considering the results of Miller (1986) and McGlynn (1990), Bassino et al. (1992, 1993) decided to investigate the issue with numerical experiments. Adopting parameters similar to those of the galaxies of the Virgo cluster, they found that: a) Dynamical friction is very small, so that captures must be almost exclusively due to tidal effects; b) King-profile galaxies are completely destroyed after a few flybys; c) While the nucleated galaxies get quite a beating too, their nuclei manage to survive (albeit strongly depleted) at least several flybys, but it is doubtful whether they can endure much longer.

Figure 1 (from Bassino et al. 1992) shows the evolution of a dwarf galaxy orbiting around a giant one $1.5 \cdot 10^5$ times more massive. The dwarf galaxy is made up of 5,000 bodies, while the giant one is represented by a fixed potential, corresponding to a singular isothermal sphere, whose center lies at the center of coordinates; the Aguilar & White (1985) code, kindly made available to us by L.A. Aguilar and slightly modified to include the effect of dynamical friction, was used for the integrations. Apocentric and pericentric distances are 250 and 25 kpc, respectively, and the orbital period is 17 time units; the time is given on each plot. The first row shows the large scale evolution, and the dispersion of the satellite along its orbit can be clearly seen. The second row shows, on smaller scale, the evolution of a nucleated galaxy, while the third row depicts the same for a King profile galaxy. A sizable part of the nucleated dwarf still survives after six revolutions, but the King profile is on its way toward disappearing.

Acknowledgements. I am very grateful to my coworkers L.P. Bassino, D.D. Carpintero, P.M. Cincotta, S.A. Cora, J.A. Núñez, A.R. Plastino, M. Rabolli, M.M. Vergne, and F.C. Wachlin for their suggestions and help, and to S.D. Abal de Rocha, M.C. Fanjul de Correbo and R.E. Martínez for their technical assistance. Special thanks go to R.H. Barbá. This work was supported with grants from the Consejo Nacional de Investigaciones Científicas y Técnicas de la República Argentina (CONICET) and from the Fundación Antorchas.

References

Aarseth S.J., 1985, Direct Methods for N-Body Simulations. In: Multiple time scales. Academic Press, New York, London, p. 377
Aguilar L.A., White S.D.M., 1985, ApJ 295, 374
Aguilar L.A., White S.D.M., 1986, ApJ 307, 97
Alladin S.M., 1965, ApJ 141, 768
Barnes J.E., Hut P., 1986, Nat 324, 446
Barnes J.E., Hernquist L., 1992, ARA&A 30, 705

Bassino L.P., Muzzio J.C., Rabolli M., 1992, VII IAU Regional Latin American Meeting
Bassino L.P., Muzzio J.C., Rabolli M., 1993, (in preparation)
Bekenstein J.D., Maoz E., 1992, ApJ 390, 79
Binney J., Tremaine S., 1987, Galactic Dynamics, Princeton Univ. Press, Princeton
Bontekoe Tj. R., van Albada T.S., 1987, MNRAS 224, 349
Carpintero D.D., Muzzio J.C., Vergne M.M., 1989, Cel. Mech. 45, 119
Cincotta P.M., Muzzio J.C., Núñez J.A., 1991, Cel. Mech. & Dyn. Astron. 52, 263
Cora S.A., Muzzio J.C., Vergne M.M., 1993, (in preparation)
Chandrasekhar S., 1943, ApJ 97, 255
Dekel A., Lecar M., Shaham J., 1980, ApJ 241, 946
Faber S.M., Jackson R.E., 1976, ApJ 204, 668
Forte J.C., Martínez R.E., Muzzio J.C., 1982, AJ 87, 1465
Freeman K.C., 1990, Our Fossil Galaxy. In: Wielen R. (ed.) Dynamics and Interactions of Galaxies. Springer-Verlag, Berlin, Heidelberg, p. 36
Fukushige T., Ito T., Makino J., Ebisuzaki T., Sugimoto D., Umemura M., 1991, PASJ 43, 841
Gelato S., Chernoff D.F., Wasserman I., 1992, ApJ 384, 15
Grillmair C.J., 1992, Thesis, The Australian National University
Gurzadyan V.G., Savvidy G.K., 1986, A&A 160, 203
Heggie D.C., Ramamani N., 1993. Approximate Self-Consistent King Models. In: Djorgovski S., Meylan G. (eds.) ASP Conf. Ser. 50, Structure and Dynamics of Globular Clusters. ASP, San Francisco, in press
Hénon M., 1958, Ann. d'Astrophys. 21, 186
Hénon M., 1960, Ann. d'Astrophys. 23, 467
Hénon M., 1966, Bull. Astron., 3e. Série, 1, 1, 57
Hernquist L., Weinberg M.D., 1989, MNRAS 238, 407
Hernquist L., Ostriker J.P., 1992, ApJ 386, 375
Innanen K.A., 1979, AJ 84, 960
Ito T., Ebisuzaki T., Makino J., Sugimoto D., 1991, PASJ 43, 547
Jefferys W.H., 1966, AJ 71, 306
Kalnajs A., 1972. Polarization Clouds and Dynamical Friction. In: Lecar M. (ed.) Gravitational N-body problem. Reidel, Dordrecht, p. 13
Kandrup H.E., 1983, Ap&SS 97, 435
King I.R., 1962, AJ 67, 471
Knobloch E., 1978, ApJS 38, 253
Landau L.D., Lifshitz E.M., 1960. Mechanics. Pergamon Press, Oxford
Leeuwin F., Combes F., Binney J., 1993, MNRAS (in press)
Lin D.N.C., Tremaine S., 1983, ApJ 264, 364
Makino J., 1991, PASJ 43, 859
McGlynn T.A., 1990, ApJ 348, 515
McGlynn T.A., Borne K.D., 1991, ApJ 372, 31
Merritt D.R., 1987, Stability of Elliptical Galaxies. Numerical Experiments. In: de Zeeuw T. (ed.) Proc. IAU Symp. 127, Structure and Dynamics of Ellipticals Galaxies, p. 315
Miller R.H., 1978, ApJ 223, 122
Miller R.H., 1986, A&A 167, 41
Miller R.H., Smith B.F., 1985, (unpublished)
Miller R.H., 1993, this volume
Muzzio J.C., 1986, ApJ 301, 23
Muzzio J.C., 1987, PASP 99, 245

Muzzio J.C., 1988, Cluster Swapping. In: Grindlay J.E., Davis Philip A.G. (eds.) Proc. IAU Symp. 126, The Harlow-Shapley Symposium on Globular Cluster Systems in Galaxies. Kluwer Academic Publishers, Dordrecht, p. 297
Muzzio J.C., Martínez R.E., Rabolli M., 1984, ApJ 285, 7
Muzzio J.C., Vergne M.M., 1988, Astrophys. Lett. & Comm. 27, 23
Muzzio J.C., Cincotta P.M., Núñez J.A., Carpintero D.D., 1988, Bol. Acad. Nac. Cien., Córdoba, Argentina, 58, 403
Muzzio J.C., Plastino A.R., 1992, Rev. Mex. Astron. Astrofís. 24, 129
Núñez J.A., Cincotta P.M., Muzzio J.C., 1993, this volume
Oh K.S., Lin D.N.C., 1992, ApJ 386, 519
Oh K.S., Lin D.N.C., Aarseth S.J., 1992, ApJ 386, 506
Okumura S.K., Ebisuzaki T., Makino J., 1991, PASJ 43, 781
Pesce E., Capuzzo-Dolcetta R., Vietri M., 1992, MNRAS 254, 466
Pfenniger D., 1986, A&A 165, 74
Rix H.W.R., White S.D.M., 1989, MNRAS 240, 941
Saslaw W.S., 1985, Gravitational Physics of Stellar and Galactic Systems. Cambridge Univ. Press, Cambridge
Smith H. Jr., 1992, ApJ 398, 519
Spitzer L. Jr., 1958, ApJ 127, 17
Sridhar S., Nityananda R., 1990, Exact Time Dependent Solutions - Violent Nonrelaxation. In: Wielen R. (ed.) Dynamics and Interactions of Galaxies. Springer-Verlag, Berlin, Heidelberg, p. 375
Toomre, A. 1977, Mergers and Some Consequences. In: Tinsley B.M., Larson R.B. (eds.) The Evolution of Galaxies and Stellar Populations. Yale Univ. Observatory, New Haven, p. 401
Tremaine S., 1981, Galaxy mergers. In: Fall S.M. and Lynden-Bell D. (eds.) The Structure and Evolution of Normal Galaxies. Cambridge Univ. Press, Cambridge, p. 67
Tremaine S., Weinberg M.D., 1984, MNRAS 209, 729
Udry S., Pfenniger D., 1988, A&A 198, 135
van den Bergh S. 1990, J.R. Astron. Soc. Can. 84, 60
Vergne M.M., 1992, Ph.D. Thesis, La Plata University
Villumsen J.V., 1982, MNRAS 199, 493
Wachlin F.C., Rybicki G.B., Muzzio J.C., 1993, MNRAS, (in press)
Weinberg M.D., 1986, ApJ 300, 93
Weinberg M.D., 1989, MNRAS 239, 549
White S.D.M., 1982, Environmental Influences on Galaxy Structure. In: Martinet L., Mayor M. (eds.) Morphology and Dynamics of Galaxies. Geneva Observatory, Geneva, p. 289
White S.D.M., 1983, ApJ 274, 53
Zamir R., 1992, ApJ 392, 65
Zaritsky D., White S.D.M., 1988, MNRAS 235, 289
Zinnecker H., Keable C.J., Dunlop J.S., Cannon R.D., Griffiths W.K., 1988, The Nuclei of Nucleated Dwarf Elliptical Galaxies - Are they Globular Clusters? In: Grindlay J.E., Davis Philip A.G. (eds.) Proc. IAU Symp. 126, The Harlow-Shapley Symposium on Globular Cluster Systems in Galaxies. Kluwer Academic Publishers, Dordrecht, 603.

How Faithful Are N-Body Simulations of Disc Galaxies? – Artificial Suppression of Stellar Dynamical Instabilities

Alessandro B. ROMEO

Onsala Space Observatory, Chalmers Univ. of Technology, S-43992 Onsala, Sweden

Abstract. High-softening two-dimensional models, frequently employed in N-body experiments, do not provide faithful simulations of real galactic discs. A prescription (♠) is given for choosing meaningful values of the softening length. In addition, a local stability criterion (♣) is given for choosing meaningful input values of the Toomre parameter for a given softening length. Such a criterion should also provide a key to a correct interpretation of computational results in terms of real phenomena.

1. Introduction

N-body simulations employing particle-mesh codes have nowadays become a very powerful tool for investigating the dynamics of disc galaxies. In particular, two-dimensional N-body models in which the stars and the cold interstellar gas are treated as two different components have successfully been applied in studies of spiral structure (e.g., Salo 1991; Thomasson 1991). A correct interpretation of computational results in terms of real phenomena poses serious problems, also because there are quantities introduced for numerical reasons which do not have clear physical counterparts. One of such artificial quantities is the softening length of the modified (non-Newtonian) gravitational interaction between the computer particles, and its value can critically affect the results of N-body experiments. It is thus of fundamental importance to have a prescription for choosing meaningful values of the softening length. From the stability point of view, it has been suggested that softening introduces a quite reasonable thickness correction for a two-dimensional model (e.g., Sellwood 1986, 1987; see also Byrd et al. 1986). In this paper the analogy between numerical softening and finite-thickness effects is investigated in detail on the basis of a local linear stability analysis, and in particular the question "How faithfully does the softening mimic the thickness of galactic discs?" is addressed. It is found that high-softening two-dimensional models, frequently employed in N-body experiments, do *not* provide faithful simulations of real galactic discs. A prescription (♠) is given for choosing meaningful values of the softening length.

Strictly connected with that problem is the choice of meaningful input values of the local stability parameter for a given softening length. In contrast to the softening length, the Toomre parameter is directly related to observable quantities, has a clear physical meaning, and its output values in N-body experiments can be compared to those predicted by theories of spiral structure and secular heating. In this paper a local stability criterion (♣) is found in virtue of the descriptive similarity between numerical softening and finite-thickness effects. Such a criterion should indeed provide a key to this problem.

A more thorough discussion is given by Romeo (1993). In this short paper we just focus on a few points.

2. Local Stability

It is convenient to adopt the following scaling and parametrization:

$$\bar{\lambda} \equiv \frac{k_{\rm H}}{|k|}, \quad \text{where} \quad k_{\rm H} \equiv \frac{\kappa^2}{2\pi G\sigma_{\rm H}}; \tag{1}$$

$$\alpha \equiv \frac{\sigma_{\rm C}}{\sigma_{\rm H}}, \quad \beta \equiv \frac{c_{\rm C}^2}{c_{\rm H}^2}; \quad Q_{\rm H} \equiv \frac{c_{\rm H}\kappa}{\pi G\sigma_{\rm H}}; \quad \eta \equiv k_{\rm H} s. \tag{2}$$

In these formulae, k is the local radial wavenumber of the perturbation, κ is the epicyclic frequency, σ_i and c_i ($i = {\rm H}, {\rm C}$) are the unperturbed surface densities and the equivalent planar acoustic speeds of the stars (H) and the cold interstellar gas (C), respectively, s is the softening length of the modified gravitational interaction. The case $\eta = 0$ represents the limit of an unsoftened gravitational interaction. There exists a critical value of the softening length beyond which the model is locally stable even for vanishing $Q_{\rm H}^2$:

$$\text{STABILITY OF COLD MODELS}: \quad s > s_{\rm crit} = \frac{1}{\rm e}\,\frac{2\pi G\sigma}{\kappa^2}, \tag{3}$$

σ being the total unperturbed surface density. This two-component extension of Miller (1972, 1974) criterion for cold models ($c_i = 0$) is indeed the limiting case of a more general *local stability criterion* for cool models ($c_i > 0$), which can be viewed as the softened two-component extension of Toomre (1964) criterion:

$$\clubsuit\ \text{STABILITY OF COOL MODELS}: \quad Q_{\rm H}^2 > \bar{Q}^2, \ \clubsuit \tag{4}$$

$\bar{Q}^2 = \bar{Q}^2(\alpha, \beta, \eta)$ being the global maximum of the marginal stability curve derived by Romeo (1993). In particular, in low-softening standard star-dominated regimes

$$\bar{Q}^2 \approx 1 + 4\,(\alpha - \eta) \quad [\alpha \ll 1;\ \beta, \eta = O(\alpha)]. \tag{5}$$

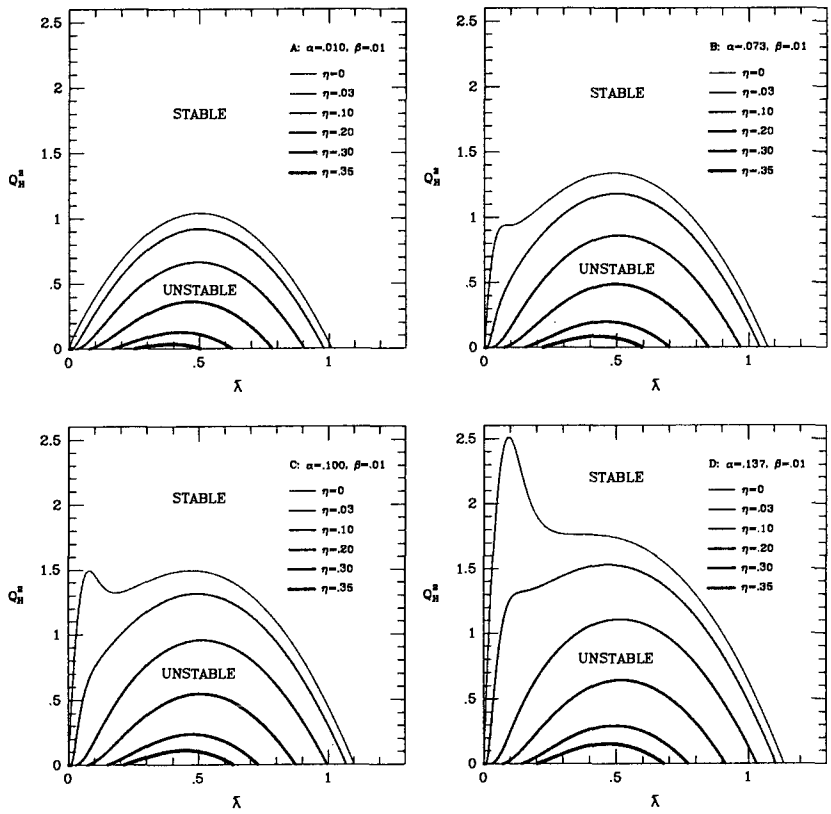


Fig. 1. Two-fluid marginal stability curves in the $(\bar{\lambda}, Q_H^2)$ plane for some values of the local parameters η, α and fixed $\beta = 0.01$. The case $\eta = 0$ represents the limit of an unsoftened gravitational interaction

3. Results

In presenting the results of the local stability analysis performed in this paper, we have considered the standard star-dominated and the peculiar gas-dominated regimes already investigated in the context of thick two-component galactic discs (Romeo 1990, 1992). The marginal stability curves shown in Fig. 1 should qualitatively be compared to those shown in Fig. 4 of Romeo (1992). It is apparent that, because of the highly stabilizing role of numerical softening, the local linear stability properties of two-dimensional N-body models can indeed be considerably different from those of thick galactic discs, as derived analytically. In particular, *note* the suppression of the gaseous peak in peculiar gas-dominated regimes even for exceedingly low softening. The softening can faithfully mimic the thickness of galactic discs or, more precisely, the effective thickness-scale of the stellar component [defined in

Eq. (6) of Romeo (1992)] only in standard star-dominated regimes, provided the softening length is chosen to be very short compared to the characteristic wavelength corresponding to the stellar peak:

$$\spadesuit \; s \ll \frac{1}{2} \frac{2\pi G \sigma_{\rm H}}{\kappa^2} \sim 1 \text{ kpc} , \; \spadesuit \tag{6}$$

as a typical value for realistic N-body models of disc galaxies. Softening lengths comparable to the critical value given by the Miller criterion do *not* fulfil this prescription.

4. Concluding Remarks

The suggestion that softening introduces a quite reasonable thickness correction for a two-dimensional model is quantitatively confirmed in low-softening standard star-dominated regimes. A constant softening length would then ideally correspond to a constant scale-height of the stellar component. On the other hand, a realistic simulation of the vertical structure of disc galaxies would in any case require a proper three-dimensional model.

Although the local stability properties of high-softening two-dimensional N-body models are considerably different from those of thick galactic discs, the propagation properties of the spiral waves are still expected to be physically plausible in standard star-dominated regimes, provided the softening length is chosen to be shorter than the critical value given by the Miller criterion [cf. the more restrictive prescription ($\spadesuit$)]. In choosing the input values of the local stability parameter as well as in comparing its output values to those predicted by theories of spiral structure and secular heating, it should then be borne in mind that the stability threshold is *not* unity (Toomre 1964 criterion for unsoftened one-component models), but the value given by the local stability criterion ($\clubsuit$) discussed in Sect. 2.

References

Byrd G.G., Valtonen M.J., Sundelius B., Valtaoja L., 1986, A&A 166, 75
Miller R.H., 1972, In: Lecar M. (ed.) Proc. IAU Colloq. 10, Gravitational N-Body Problem. Reidel, Dordrecht, p. 213
Miller R.H., 1974, ApJ 190, 539
Romeo A.B., 1990, PhD thesis, SISSA, Trieste, Italy
Romeo A.B., 1992, MNRAS 256, 307
Romeo A.B., 1993, A&A (submitted)
Salo H., 1991, A&A 243, 118
Sellwood J.A., 1986, In: Hut P., McMillan S.L.W. (eds.) The Use of Supercomputers in Stellar Dynamics. Springer-Verlag, Berlin, p. 5
Sellwood J.A., 1987, ARA&A 25, 151
Thomasson M., 1991, PhD thesis, Chalmers Univ. of Technology, Göteborg, Sweden
Toomre A., 1964, ApJ 139, 1217

SPH Simulations of the Gas Flow in Normal Spiral Galaxies

P.A. PATSIS [1] & N. HIOTELIS [2]

[1] ESO, Karl Schwarzschild Str. 2, D-8046, Garching bei München, Germany
[2] Department of Astronomy, University of Athens, GR-157 83, Hellas

1. Introduction

The value of the pattern speed (Ω_p) of density waves in spiral galaxies is usually estimated by associating major resonances (e.g., 4/1, corotation or −2/1) with spiral features such as the ends of the arms (Contopoulos & Grosbøl 1988; Patsis et al. 1991, Elmegreen et al. 1992). Choosing in this way a value for the pattern speed of a particular normal spiral galaxy, we may follow the response of stellar or gaseous models for this galaxy to an imposed spiral potential. Thus we can estimate whether our model responds better to a rapidly or to a slowly rotating pattern. Also, by comparing stellar and gaseous models with the same Ω_p for the same galaxy, we can trace differences in the dynamical behaviour of the two components. In this paper we present some gaseous models for normal grand-design spiral galaxies, and we outline the differences between them and the corresponding stellar cases.

2. The Method

For our calculations we use a version of the SPH method with a spatially and temporally varying resolution length h. For details of the code the reader may refer to Hiotelis & Voglis (1991), and references therein. The imposed potential consists of two parts, an axisymmetric part (V_0) and a spiral perturbation (V_1). V_0 is obtained from the circular rotational velocity

$$v = v_{\max} \sqrt{f\, b\, r \exp(-b\, r) + [1 - \exp(-d\, r)]}\,, \tag{1}$$

where f, b, and d are constants. The rotation curve (1) starts at $v = 0$ for $r = 0$ and tends to $v_{\max}$ as r tends to infinity. The parameters $v_{\max}$, f, b, and d are obtained by fitting Eq. (1) to existing observational data for the rotation curve and making the appropriate corrections for projection effects. The spiral perturbation is of the form:

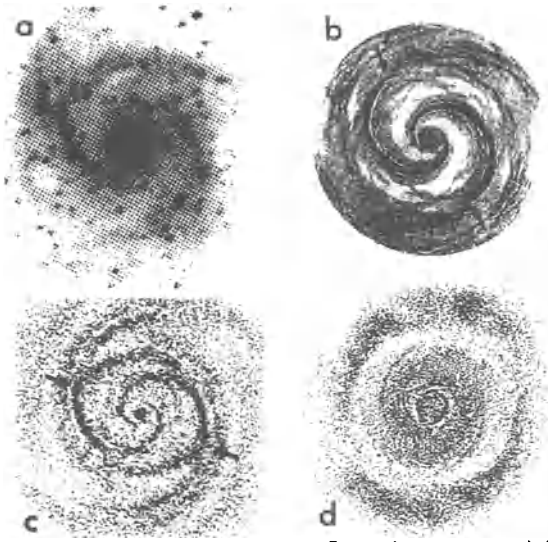


Fig. 1. A deprojected red image of NGC 2997 (a), and three non-linear models for this galaxy. In (b) and (c) the end of the spirals is assumed at the 4/1 resonance. (b) is a stellar, while (c) a gaseous model. In (d) we give a gaseous model that places the end of the spirals at the $-2/1$ resonance

$$V_1(r) = A\, r\, \exp(-\epsilon_s\, r) \left[\cos\left(2\frac{\ln r}{\tan i} - 2\theta\right)\right] , \qquad (2)$$

where ϵ_s is the inverse scale length of the spiral, and A its amplitude.

3. Results and Conclusions

The main conclusion of this study is that our results depend directly on the type of galaxy we are trying to model. In particular we can say:

- *In open, late-type, grand-design, normal spiral galaxies we have the best response in models with strong, slowly rotating patterns, which place the 4/1 resonance at the end of the observed spirals.*
- *In tightly wound, early-type, grand-design, normal spiral galaxies we have best response in models with weak, rapidly rotating patterns.*

The above conclusions agree with the results of stellar models for the same types of galaxies (Contopoulos and Grosbøl 1988, Patsis et al. 1991). They also agree with the analysis of the azimuthal profiles of the relative intensities (Grosbøl 1993), which indicate the non-linearity of the perturbation in open normal spirals.

Fig. 1 shows the response of three models for the late type Sc galaxy NGC 2997. Fig. 1a is a deprojected image of the galaxy in red light. In (b) we give a stellar model, constructed by integrating a large number of non-periodic orbits (Patsis et al. 1991). The end of the spiral is assumed to be at the 4/1 resonance, as is also assumed for (c), which is a gaseous model. In both cases the position of the 4/1 resonance is indicated with arrows. In (d) we give the response of the gas to a rapidly rotating spiral potential. In this case the end of the spirals is assumed at the $-2/1$ resonance. The $-2/1$

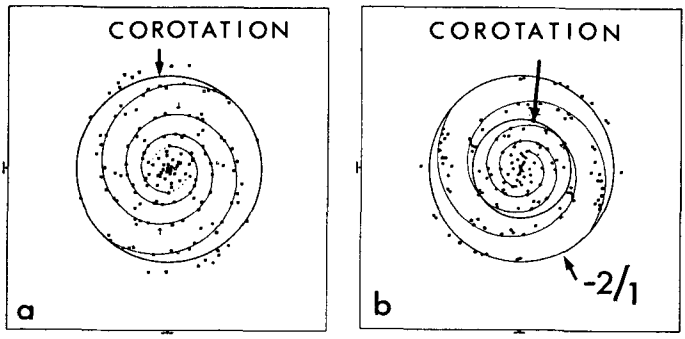


Fig. 2. The density maxima of the 2θ component as a function of radius for a corotation model (a) and for a $-2/1$ model (b).

region in this model is characterized by the presence of a ring at the outer edge of the disk. This is a typical feature found almost in all the $-2/1$ gaseous models we examined. The perturbing force is 16% of the axisymmetric one at the end of the spirals for the stellar case and about 13% for the gaseous models. Detailed comparisons between models with larger and smaller values of Ω_p can be found in Patsis et al. (1991, 1993). Here we want to emphasize that the presence of pressure and viscosity enhances secondary features in the region of the 4/1 resonance like the bifurcations of the arms observed in (c). We identify them with main interarm features of the galaxies, which are more pronounced at blue images.

In Fig. 2 we give the density maxima of the 2θ component of two rapidly rotating disks for the tight-wound spiral NGC 1357 as a function of radius. This is an early type Sa galaxy with pitch angle $i = 8°$. In (a) the end of the observed spirals corresponds to corotation, while in (b) it is at the $-2/1$ resonance. In both cases, a weak perturbation of the order of 1% is able to give a nice response over the whole region of the disk, without problems at the 4/1 region. The overplotted spirals represent the imposed potential minima. The corotation model can be regarded as better, since response maxima of the 2θ component do not tend to form a spiral tighter than the imposed one in the outer parts. Also, in general, we observe that the points at the corotation model fit the potential minima better all over the disk.

References

Contopoulos G., Grosbøl P., 1988, A&A 197, 83
Elmegreen B.G., Elmegreen D., Montenegro L., 1992, ApJS 79, 37
Grosbøl P., 1993, P.A.S.P. in press
Hiotelis N., Voglis N., 1991, A&A 243, 333
Patsis P.A., Contopoulos G., Grosbøl P., 1991, A&A 243,373.
Patsis P.A., Hiotelis N., Contopoulos G., Grosbøl P., 1993, submitted to A&A

Regular Orbits and Cantori in the Potential of the Barred Galaxy NGC 936

Hervé WOZNIAK [1 2]

[1] Observatoire de Genève, CH-1290 Sauverny, Switzerland
[2] Observatoire de Marseille, 2, Pl. Le Verrier, F-13248 Marseille cédex 4, France

Traditionally, the potential used for the study of the orbital structure of a bar is derived from a more or less realistic analytical mass distribution. Our knowledge of families of periodic orbits, bifurcations and ergodicity in bars and lenses is thus based on potentials generally computed at the best from Ferrer's ellipsoids. However, the light distribution of real galaxies does not look like the mass distribution of the models: it has been shown (Athanassoula et al. 1990) that light distributions of strong bars are rectangular like and not elliptical. In order to get a more realistic expression for the potential, we have computed it directly from CCD frames following a method briefly described in Wozniak (1991). Assuming an infinite thinness for a galaxy and a constant mass-to-light ratio, we can solve numerically the 2D Poisson equation for a given galaxy using the projected light distribution as a model for the surface density of mass. We can neglect any massive dark halo since it is expected to have a very little influence on the dynamics inside the bulge and the bar. We have applied this method to a sample of 17 early-type barred galaxies (Athanassoula & Wozniak, 1993). Here, we only deal with one of them, NGC 936, for which we have computed orbits in the calculated potential.

Our calculations show that the ellipse-like family (x_1) is the backbone of the bar and it has been shown in Athanassoula et al. (1990) that the outline of bars is rectangle-like. Thus we would like to discuss orbits that could in principle support this shape. We will only describe some types of orbits that are good candidates. Our arguments are only based on the shape of the orbits and the time spent along their different section. For more details, see Wozniak & Athanassoula (1993).

Trapped orbits. A traditional way to make the bar rectangular is to consider the trapped orbits around the x_1 family. The major axis of these orbits oscillates around the bar major axis and thus an area with a rectangular aspect could be filled although the x_1 orbits themselves are elliptical or even pointed at the major axis (Figs. 1d–f). Indeed, if we superpose some of these trapped orbits taken at different values of E_J, we may reproduce the rectangular shape of the bar but only in the inner region where the isodensities contours are thin and their semi-major axis shorter than the bar length

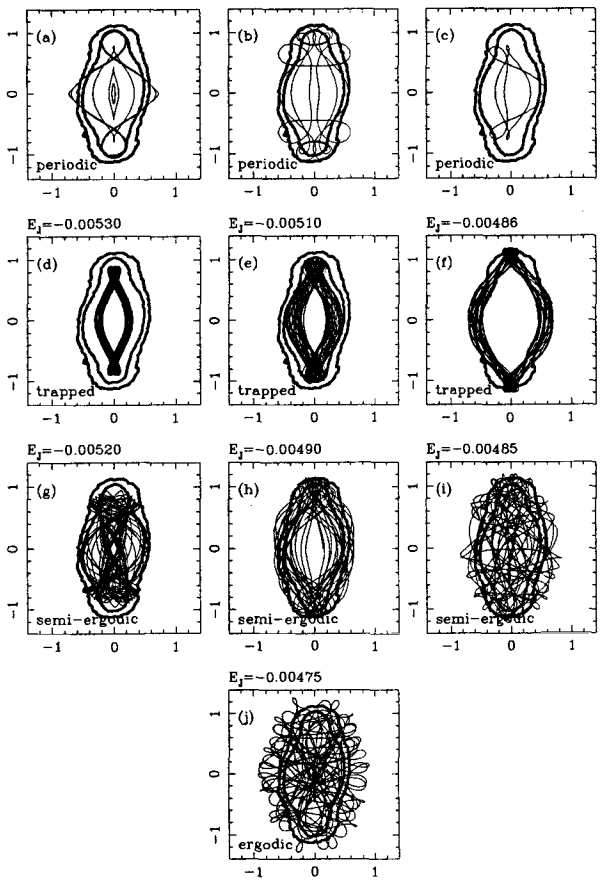


Fig. 1.
a-c: Examples of periodic orbits, resp. x_1, 4/1 and 3/1 orbits. The two thick lines represent two isophotes of the bar of NGC 936, the outer one being that of highest rectangularity.
d-f: Examples of trapped orbits. These orbits were computed for $27 \cdot 10^8$ yr.
g-i: Examples of semi-ergodic orbits.
j: Example of an ergodic orbit

(cf. Figs. 1d–f). Thus there are no trapped orbits that can reproduce the observed width of the bar where its rectangularity is maximum.

Semi-chaotic orbits. We can also consider semi-trapped (or semi-chaotic) orbits, that are orbits confined in a cantorus surrounding the x_1. They could participate in the building of a rectangular bar, as for the self-consistent models of Pfenniger (1984) or the N-body bar of Sparke & Sellwood (1987). These orbits appear at $E_J \approx -0.0052$ (cf. Fig. 1g). They semi-chaotic orbits have the same extent along the major-axis as the trapped orbits at the same E_J but they fill a wider area. Thus, at $E_J \approx -0.0049$ semi-chaotic orbits have the right size to fill the area defined by the outer isophote (cf. Fig. 1h). Furthermore these orbits spend much of their time in the "corners" of the bar, as we have checked by computing the corresponding orbital density. The domain in E_J for which semi-chaotic orbits can participate in the rectangularity of the bar is small (cf. Fig. 2).

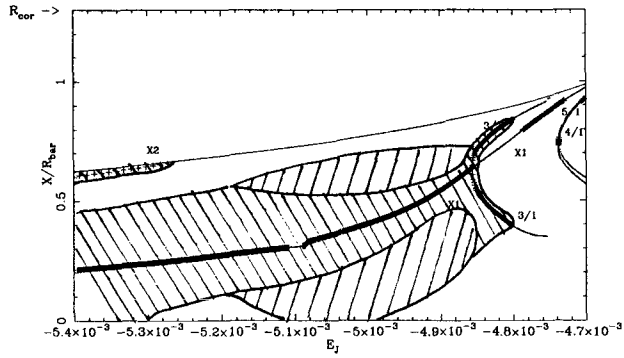


Fig. 2. Characteristic diagram: blow-up of the energy region where trapped orbits and semi-ergodic orbits are responsible for the rectangle-like shape of the bar. E_J is the Jacobi constant, X is the intersection of a periodic orbit with the semi-minor axis of the bar. Stable orbits are plotted with a + and unstable ones with a dot. The thin solid line is the Zero Velocity Curve. The \ shaded regions give the domain of orbits trapped around a family of periodic orbits; the / one is filled by semi-ergodic orbits. The remaining areas are essentially covered by ergodic orbits

Discussion. The lack of stable families of periodic orbits over part of the extent of the bar need not be destructive for the bar. Indeed, there are always semi-chaotic orbits which could have a non negligible weight in the distribution function of the dynamical system. Moreover, it has been shown by Pfenniger (1984) that a substantial amount of chaotic or semi-chaotic orbits could be found in a self-consistent model even if models free of semi-chaotic orbits can be built. Petrou (1984) reached similar results without building a self-consistent model. In the models of Pfenniger, semi-chaotic orbits are essentially found near the unstable Lagrangian points $L_{1,2}$ around which, in our model, the bar isophotes are the most rectangular.

Obviously, this is no substitute for the proper calculation of a self-consistent model using, for instance, the Schwarzschild (1979) method. Such a method could give us the distribution function and thus full information on what kind of orbits participate in the building of the bar. This will be done in a future paper.

References

Athanassoula E., Morin S., Wozniak H., Puy D., Pierce M.J., Lombard J., Bosma A., 1990, MNRAS 245, 130
Athanassoula E., Wozniak H., 1993, preprint
Petrou M., 1984, MNRAS 211, 283
Pfenniger D., 1984, A&A 141, 171
Schwarzschild M., 1979, ApJ 232, 236
Sparke L.S., Sellwood J.A., 1987, MNRAS 225, 653
Wozniak H., 1991, in: Dynamics of Disc Galaxies, ed. B. Sundelius, p. 175
Wozniak H., Athanassoula E., 1993, A&A submitted

The Role of Stochastic Motion in a Central Field with a Bar-Like Perturbation

Josué A. NÚÑEZ, Pablo M. CINCOTTA, & Juan C. MUZZIO

Facultad de Ciencias Astronómicas y Geofísicas, UNLP
PROFOEG - CONICET

Abstract. The motion of a particle in a central field with a rotating bar-like perturbation has been investigated by many authors, but, only a few papers deal with fixed bar perturbations which arise, for example, as a result of the radial orbit instability (e.g. D. Merritt & L.A. Aguilar 1985, MNRAS 217, 787). Our main interest is to investigate the relevance of stochastic orbits in such kind of fields. Thus, using information theory, we develop a simple method to evaluate the global degree of stochasticity of a set of orbits in a given Hamiltonian. Briefly, we obtain the Poincaré Surface of Section (PSS) for each orbit, then we choose a partition of it and, finally, we compute the entropy of the PSS defined as $S = -\sum_{i=1}^{N} P_i \ln P_i$, where P_i is the probability of finding an intersection of the orbit with the PSS within the i^{th} partition. We expect S to have two regimes: a low-value one for regular orbits and a high-value one for stochastic orbits. To test our method, we used the well-known Hénon-Heiles potential. We obtained the PSS with 2500 points for 70 orbits. Our results showed the power of the method to describe the features of the potential, the lower values of S corresponding to stability islands and the higher ones to clearly stochastic regions.

We then applies this algorithm to the isochrone potential, $\phi_I(r)$, plus a bar-like perturbation of the form $\phi_1(r,\varphi) = -\alpha\psi(r)\sin 2\varphi$, where α is the perturbation paramenter and a is a constant. We consider 90 orbits, all of them with low angular momenta. Our results show that about 50% of the orbits are stochastic, although the effective perturbation ϕ_1/ϕ_I is not larger than 4%. This result is very interesting since, as the probability of finding a system with perfect spherical (or axial) symmetry is very low, there is a good chance of finding stochastic orbits for the low-angular-momentum stars that move within such a field.

Acknowledgements. The authors are very grateful to Lic. D.D. Carpintero for his advice on the use of computer software and codes, and to M.C. Fanjul de Correbo, S. Abal de Rocha, and R.E. Martínez for they technical support. This work was supported by grants from the Fundación Antorchas and from the Consejo Nacional de Investigaciones Científicas y Técnicas de la República Argentina, whose assistance is gratefully acknowledged.

A Hierarchical Model of Patchy-Structured Galaxies and Evolutionary Processes

E.M. NEZHINSKIJ [1] & A. OLLONGREN [2]

[1] Inst. for Theoretical Astronomy, Naberezhnaya Kutuzova 10, 191187 St. Petersburg, Russia
[2] Dept. of Computer Science and Astron. Dept., Leiden Univ., Niels Bohrweg 1, 2333 CA Leiden, the Netherlands

Abstract. Results of an investigation concerned with the development of hierarchical models of S- and SB-galaxies are presented.

In (Nezhinskij & Ollongren 1992) new dynamical models of S- and SB-galaxies were proposed. In these models *stellar complexes* (as Efremov 1984 called giant condensations of matter in the plane of real spiral galaxies) are considered to be point bodies situated in the plane of symmetry forming stable central configurations (i.e., self-gravitational systems of bodies rotating around the center of mass of the model with constant angular velocity). In S- and SB-galaxies these configurations can be seen as patchy rings and ring- like patterns.

The present paper is based on an extension of the mentioned models achieved by constructing *hierarchical* representations of S- and SB-galaxies. The foundations for this choice are well-known facts: the hierarchical structure of quasi-stationary galaxies and the occurrence of the mentioned ring-like spiral patterns in these galaxies. The complexes forming these patterns consist of gigantic clouds of gas and dust, open stellar clusters and associations. Therefore *globules* of matter were chosen instead of point bodies to represent these objects. The results are summarized below.

Sufficient conditions for the existence and stability of hierarchical models have been formulated (to be published separately).

The general question on the self-gravitation of stellar complexes embedded in a galactic gravitational field of force is open. We considered models of S- and SB-galaxies which contain non-self-gravitational complexes.

To complete the discussion on the question how well the proposed hierarchical model represents an actual galactic system, the half-value time for the destruction of stellar complexes has been estimated. Using a realistic assumption on the density of matter, it turns out to be larger than 10^{10} years.

Finally several representative examples of stable models of the hierarchical type are shown in Fig. 1.

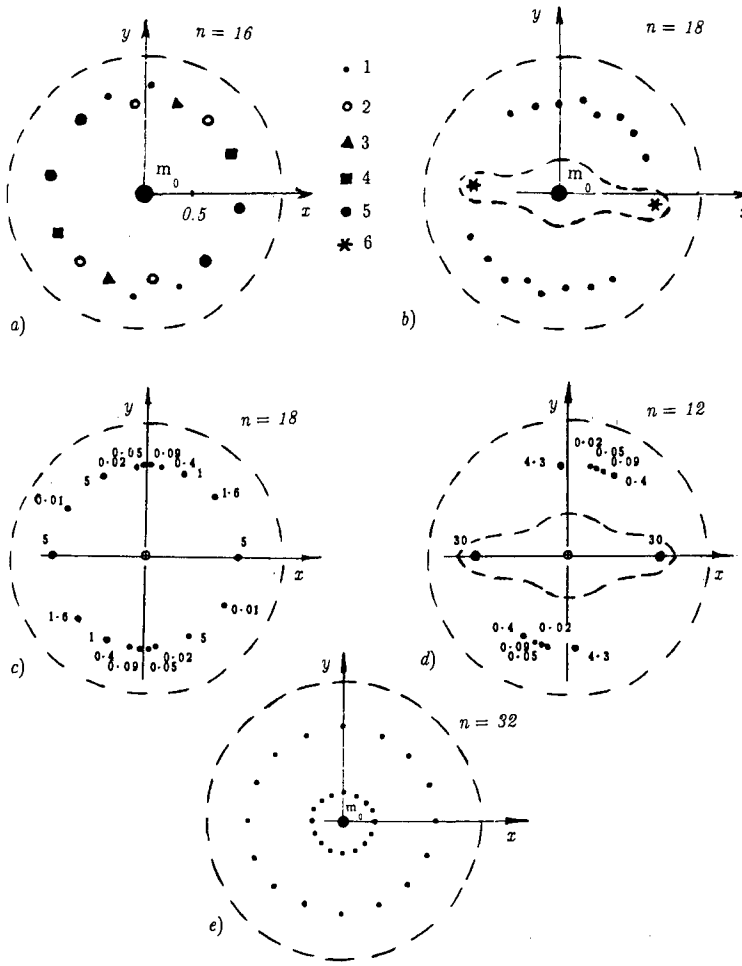


Fig. 1. The parameters for the models are chosen in such a way that the total mass of the nucleus and the part of the spherical component in the ring-like patchy pattern of globules is equal to 10^{10} $M_\odot$ and the mean radius of a patchy pattern is 8 kpc. In models a) and b) m_0 is equal to 100 in relative units, the density of the spherical component ρ is equal to 1000 and the positions of the bodies with masses $m_i = 1, 2, 3, 4, 5, 30$ are marked with the numbers $i = 1, 2, 3, 4, 5, 6$. In models c) and d) the potential function of the spherical component is $U(r) = C \ln(1 + r^2/a^2)$, with C and a constants. Close to the position of each globule its mass is written (the unit of mass is 10^7 $M_\odot$). Case e) is an example of a many-tiered model

References

Efremov Yu.N., 1984, Vestnik Akad. Nauk SSSR, No. 12, 56
Nezhinskij E.M., Ollongren A., 1992, A&A 266, 140

Evolution of Clusters of Galaxies

Yoko FUNATO [1 3], Junichiro MAKINO [2], & Toshikazu EBISUZAKI [1 3]

[1] Department of Earth Science and Astronomy,
[2] Department of Information Science and Graphics,
[3] College of Arts and Sciences,
The University of Tokyo, Meguro-ku, Tokyo 153, Japan

Abstract. We investigated the evolution of clusters of galaxies using direct N-body simulations in which each galaxy is modeled by many particles. We found that the galaxies in the inner region of cluster lose their mass more rapidly than do those in the outer region. Therefore the galaxies in the inner region become smaller than those in the outer region. Significant fraction of the total mass of the cluster escapes from galaxies into the intracluster space. They form diffuse halo around the center of cluster. We also found that the mass (m) and the velocity dispersion (σ) satisfy the relation, $m \propto \sigma^4$. This tendency is consistent with the Faber-Jackson relation, which $L \propto \sigma^4$. This implies that the Faber-Jackson relation is result of the evolution of the galaxies driven by interactions with other galaxies and the tidal field of the parent cluster.

1. Introduction

The interaction between galaxies is highly inelastic. Through encounters, the energy of orbital motion is converted to the random energy of stars within galaxies. As a result, each galaxy loses its mass and changes its structure.

The rates of change of mass and energy strongly depend on the structure of each galaxy, while the structure of each galaxy changes through the change of mass and energy. It is, therefore, necessary to perform self-consistent simulation in which each galaxy is modeled as an N-body system.

We performed a series of N-body simulations to investigate the evolution of clusters of galaxies. More detailed description of our result is given in Funato et al. (1993). In our work, each galaxy is represented by many particles and we calculate direct summation of forces to each particle. Thus, our result is not affected by any ambiguity in the assumptions on the nature of the interaction between galaxies. We used GRAPE-3; a special-purpose computer for gravitational N-body problem (Okumura et al. 1993) for the time integration.

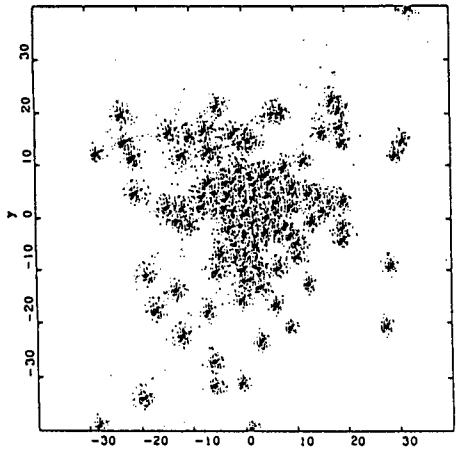


Fig. 1. Snapshot of the initial condition projected onto the x-y plane (Funato et al. 1993)

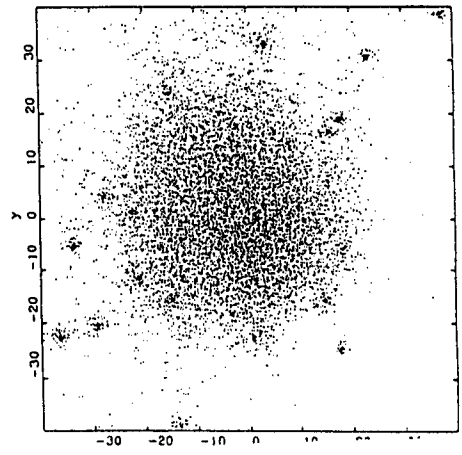


Fig. 2. Right: Snapshots of the cluster at $t = 200$ projected onto the x-y plane (Funato et al. 1993)

We found 1) galaxies in the inner region lose their mass more rapidly than do those in the outer region. Therefore, galaxies in the inner region become smaller than those in the outer region. 2) A large fraction of total mass of the cluster escapes from galaxies into the intracluster space to form diffuse halo around the center of the cluster. 3) Between the mass of each galaxy (m) and velocity dispersion of particles within the galaxy (σ), the relation known as Faber-Jackson relation $-\sigma \propto m^{1/n}, n \simeq 4-$ (Faber & Jackson 1976), develops as the cluster evolves.

2. Cluster Models

Each galaxy is represented by a Plummer model with 512 equal mass particles. The cluster comprised 128 equal galaxies. Figure 1 shows an initial cluster model. The entire cluster is represented by that of a Plummer model in the dynamical equilibrium. We used a standard units where $G = 1$, $m = 1$, and $e = -0.25$ (Heggie & Mathieu 1986). The crossing times of a galaxy and the cluster are 2.8 and 22.3, respectively. The velocity dispersion of the cluster is three times larger than that of a galaxy.

3. Results

Figure 2 is a snapshot at $t = 200$. There is a diffuse halo around the center of the cluster made of escapers and there is one massive galaxy. Here we call particles which escaped from its parent galaxy as escapers. Half of total mass of the cluster has escaped from galaxies.

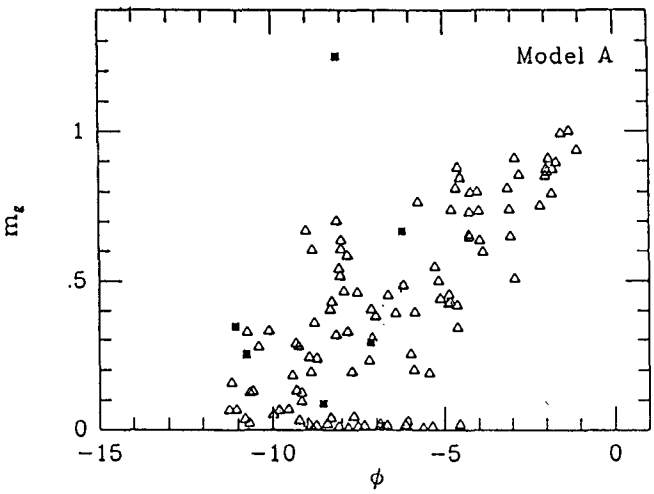


Fig. 3. Masses of the galaxies plotted against the depth of the cluster potential. Filled squares correspond to merger remnants (Funato et al. 1993)

Figure 3 shows the relation between masses of galaxies and the depth of the cluster potential at their positions. It is clear that galaxies in the bottom of the potential well (center of the cluster) are less massive than those in the outer region, at though several massive galaxies are formed through mergings. This tendency is explained by the fact that encounters are more frequent in

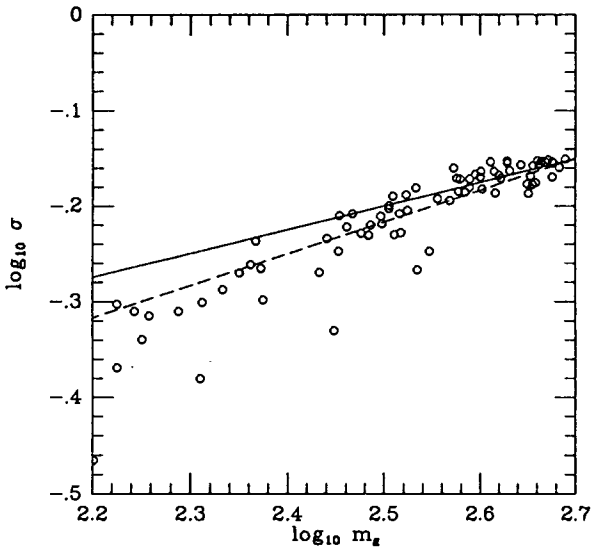


Fig. 4. Masses of galaxies plotted against their velocity dispersions. Solid and dashed lines correspond to the relation $\sigma \propto m^{1/4}$ and $\sigma \propto m^{1/3}$, respectively

the central region. This correlation is in good agreement with observation of Coma cluster by Ström & Ström (1978).

Figure 4 shows that galaxies in a cluster evolve so as to satisfy the Faber-Jackson relation. This implies that the Faber-Jackson relation is result of the interactions between galaxies.

References

Faber S.M., Jackson R.E., 1976, ApJ 204, 668
Funato Y., Makino J., Ebisuzaki T., 1993, PASJ 45, in press
Heggie D.C., Mathieu R.D., 1986, in: The Use of Supercomputers in Stellar Dynamics, P. Hut, S. Mc Millan (eds.), Springer, Berlin, p. 233
Okumura K.O., Makino J., Ebisuzaki T., Fukushige T., Ito T., Sugimoto D., Hashimoto E., Tomoda K., Miyakawa N., 1993, PASJ 45, in press
Ström K.M., Ström S.E., 1978, AJ 83, 73

Smoothing of the Cosmic Background Radiation by Multiple Gravitational Scattering

Junichiro MAKINO

Department of Information Science and Graphics, College of Arts and Sciences, University of Tokyo, Komaba 3-8-1, Meguro-ku, Tokyo 153, Japan

Abstract. We investigated the smoothing of the cosmic background radiation (CBR) through gravitational scatterings. The CBR is gravitationally scattered by galaxies, clusters of galaxies, and superclusters during the travel from the last scattering surface. Although the effect of the gravitational scatterings was thought to be unimportant, we found that the scatterings by superclusters reduce the anisotropy by a large factor. This result is explained by the fact that the distance between two light rays increases exponentially through multiple scatterings. This exponential growth occurs if the distance is smaller than the critical distance determined by the number density of the scattering objects. We found that the gravitational scatterings by superclusters can reduce the temperature anisotropy of the CBR at present time by a large factor, in angular scales up to a few degrees.

1. Introduction

The exponential instability of the orbit in N-body systems has been well known in the community of the stellar dynamics (Miller 1964; Goodman et al. 1993 and references therein). Because of the exponential instability, two N-body systems with slightly different initial conditions depart from each other exponentially in time.

The basic mechanism of this exponential instability is very simple. Consider the orbit of two test particles in the fixed distribution of point masses. When the test particles are scattered by a point mass, the distance between two particles is amplified because the orbit of the particle closer to the scatterer is deflected by a larger angle. When they are scattered at the next time, the distance is again amplified. Since the scattering angle is proportional to the impact parameter, the difference in the scattering angle is also proportional to the distance between two orbits. Therefore the distance between two orbits grows exponentially in time as long as the scattering is coherent. Figure 1 illustrates consecutive scatterings.

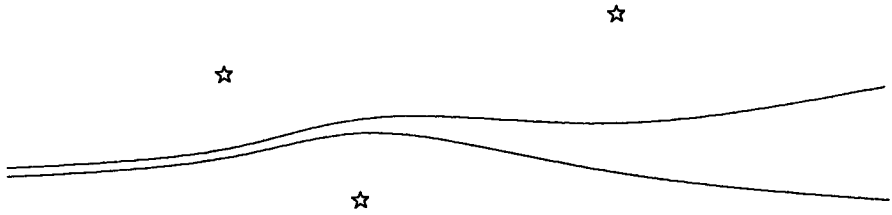


Fig. 1. Exponential divergence of two nearby orbits

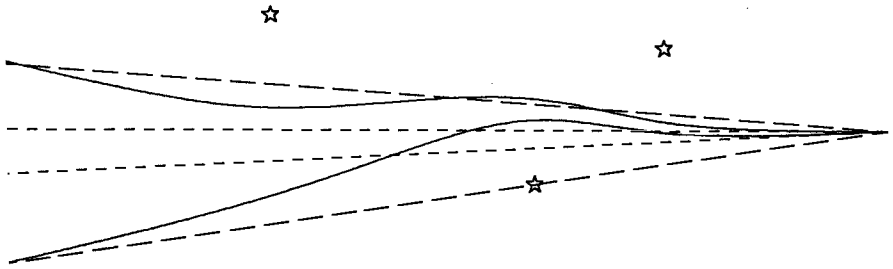


Fig. 2. The trajectories of photons from the last scattering surface to an observer

According to Goodman et al. (1993), the e-folding time of this exponential instability is of the order of 0.1 crossing time, and the exponential growth would saturate when the distance between two orbits becomes $RN^{-1/2}$, where R is the characteristic radius of the system and N is the number of particles in the system.

In this paper, we discuss the effect of this exponential instability on the orbits of photons, in particular, the observed temperature anisotropy of the cosmic background radiation (CBR). Figure 2 illustrates the observation of the CBR through scattering objects. The observer is at the right hand side of the figure and see two rays. If photons traveled following straight lines, their orbits would be the straight lines (short-dashed lines). However, actual orbits are deflected by objects such as galaxies and clusters, as shown in the solid curves. As a result, the actual beam width, shown in the long-dashed lines, is larger than the beam width at the position of the observer.

The possibility that the gravitational scatterings change the anisotropy of CBR has been discussed by many authors (e.g. Blanchard & Schneider 1987; Cole & Efstathiou 1989; Sasaki 1989; Tomita & Watanabe 1989). However, most authors assumed that the effect of the multiple scatterings could be expressed as the superposition of single scatterings. Thus their estimates did not take into account the exponential instability. In the following, I briefly describe the effect of the exponential instability on the anisotropy of CBR due to gravitational scatterings. A more detailed discussion is presented by Fukushige et al. (1993)

2. Theory

In the following, we consider how the distance between two rays emitted from the observer (see Fig. 2) changes when we calculate the trajectories backward in time. For simplicity, we assume that the universe is flat ($\Omega = 1$).

Goodman et al. (1993) gave an intuitive explanation for the mechanism of the exponential growth. If we consider the encounters with the impact parameter p, the average time interval of such encounters is given by

$$\tau \sim n^{-1} p^{-2} c^{-1} , \tag{1}$$

where n is the number density of the scattering objects and c is the light velocity. The change of the distance d between two neighboring photons is given by

$$\Delta d = \theta \tau . \tag{2}$$

Here, θ is the difference of the directions of two photons generated by an encounter and is approximately expressed as $\theta \sim Gmdp^{-2}c$, where m is the mass of the scattering objects. Note that whether the distance increases or not depends on the angle between the impact parameter vector and the relative position vector of two photons. In order for the distance to grow exponentially, Δd must be comparable to or larger than d, since if $\Delta d \ll d$, the effects of encounters with different angles tend to cancel each other to give only the random walk effect on the distance. The condition $\Delta d \simeq d$ gives

$$\tau \sim 1/\sqrt{G\rho} . \tag{3}$$

The impact parameter corresponding to $\Delta d \simeq d$ is

$$p_{\rm crit} \sim (Gm)^{1/4} n^{-1/4} c^{-1/2} . \tag{4}$$

For a flat universe with radius R, this value corresponds to

$$p_{\rm crit} = RN^{-1/2} , \tag{5}$$

where N is the total number of the scattering objects in the universe. As shown in Eq. (3), the time-scale of the growth is of the order of the free fall time, irrespective of the mass of the scattering objects.

Fukushige et al. (1993) performed a series of numerical simulations of the orbits of two photons which are initially in small distance. They found that the e-folding time of the distance between the trajectory of two photons is given by

$$\tau = 0.22/\sqrt{G\rho} , \tag{6}$$

where G is the gravitational constant and ρ is the mass density of the scattering objects. They confirmed that the time-scale is given by Eq. (6) and

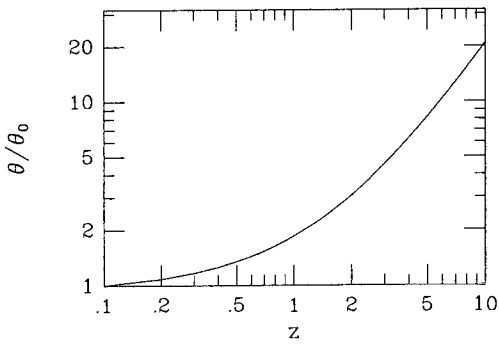


Fig. 3. The magnification factor of beam angle θ/θ_0 as a function of z

does not depends on the mass of scattering objects, if the mass density is kept constant.

The width w of the beam with angle θ_0 at $z = 0$ is given by

$$\mathrm{d}w/\mathrm{d}t = -c(\theta_0 + w/\tau)\,, \qquad w|_{z=0} = 0\,, \tag{7}$$

where t is the time and c is the light velocity. By solving Eq. (7), we obtain

$$\theta = w/ct = 0.49\,\theta_0\left[(1+z)^{1.6} - (1+z)^{-3/2}\right] / \left[1 - (1+z)^{-3/2}\right] \tag{8}$$

for the effective beam angle. Figure 3 shows θ/θ_0 as a function of z. If scattering objects such as galaxies and superclusters are formed at, say, $z = 4$, the effective beam angle to observe anything in the region $z > 4$ is larger than the beam angle at observer by a factor of 7. The actual effect on the measured anisotropy depends on the power spectrum of the anisotropy, but is likely to be significant.

The exponential growth does not occur if the universe is completely uniform. The scattering objects must be compact enough to be effective. As discussed in Sect. 2, encounters with the impact parameter larger than p_{crit} do not lead to the exponential growth. Thus, if the typical size of the scattering objects is larger than p_{crit}, the exponential growth cannot occur. In addition, the exponential growth of the distance between two rays saturates around $w \sim p_{\mathrm{crit}}$. This behavior is consistent with the result of Goodman et al. (1993). We found that galaxies are effective up to few arcseconds and superclusters are effective up to a few degrees.

References

Blanchard A., Schneider J., 1978, A&A 184, 1
Cole S., Efstathiou G., 1989, MNRAS 239, 195
Fukushige, T., Makino J., Ebisuzaki T., 1993, submitted to ApJL
Goodman J., Heggie D.C., Hut P., 1993, ApJ (in press)
Miller R.H., 1964, ApJ 140, 250
Sasaki M., 1989, MNRAS 240, 415
Tomita K., Watanabe K., 1989, Prog. Theor. Phys. 82, 563

Angular Momentum of Galaxies Within the Local Supercluster

Włodzimierz GODŁOWSKI

Astronomical Observatory, Jagiellonian University, ul. Orla 171, 30-244 Kraków, Poland

Abstract. The various available scenarios for the galaxy origin predict different types of galactic angular momenta distributions within galactic superclusters. The random distribution of galaxy rotation axes is expected to occur within the hierarchical clustering picture (Dekel 1985), whereas some alignment is expected to arise in the fragmentation scenario. Within the latter model the perpendicular to the supercluster main plane orientation of galactic rotation axes conforms to the predictions of the turbulence model (Efstathiou & Silk 1983), while the alignment of galactic axes with that plane supports rather the adiabatic galaxy origin model (Shandarin 1974). The recently proposed "hedgehog model" predicts that the the galaxy planes should tend to be oriented perpendicular to direction oriented toward the supercluster centre. Most of the papers published to date analyze the Local Supercluster (e.g. Hawley & Peebles 1975; Kapranidis & Sullivan 1983; MacGillivray & Dood 1985) to conclude that galactic rotation axes are oriented in a random manner or, when some alignment was found, the rotation axes were preferably oriented perpendicularly to the supergalactic plane. In those papers however, the samples considered were selected according to magnitude limits so they could be contaminated by background objects. Moreover, using the equatorial coordinate system made the discussion of results very difficult. In this paper a method originally proposed by Jaaniste & Saar (1977) for investigating galactic angular momentum in the Local Supercluster (LSC) is applied to a sample of 2227 galaxies chosen according radial velocities from the UGC and ESO catalogues. We (Flin & Godlowski 1986; Godlowski 1993) introduced the important modifications and corrected for its inconsistencies and use sophisticated statistical method which had not been used in solving this problem before using supergalactic coordinate system. The method takes into account both the galactic image position angle and the inclination of the normal to the galaxy to the line-of sight. Another independent sample of galaxies taken from Tully's Nearby Galaxy Catalogue is also used. In the present paper, we find the distribution of galactic planes throughout LSC anisotropic. The planes tend to be oriented perpendicular to the Local Supercluster plane. The projection of galactic rotation axes on the LSC plane

shows a tendency to point toward the Virgo Cluster centre. Differences between spiral and non-spiral galaxies are observed, with a weaker alignment for spirals. When we check our computation using independent sample taken from Tully's catalogue we obtain very similar results. Our results support the so called "pancake" galaxy formation scenario, but the "hedgehog model" is not excluded either because of differences distribution of edge-on galaxies.

References

Dekel A., 1985, ApJ 298, 461
Efstathiou G.A., Silk J., 1983, Formation of Galaxies, Fundam. Cosmic Phys. 9, 1
Flin P., Godlowski W., 1986, MNRAS 222, 525
Godlowski W., 1993, MNRAS, accepted
Hawley D.L., Peebles P.J.E., 1975, AJ 80, 477
Jaaniste J., Saar E., 1977, Tartu Obs. Preprint A-2
Kapranidis S., Sullivan III W.T., 1983, A&A 118, 33
MacGillivray H.T., Dood R.J., 1985, A&A 145, 269
Shandarin S.F., 1974, Sov. Astron. 18, 392

8. Concluding Remarks

10 Problems the Solutions of Which Can Seriously Influence Stellar Dynamics

V.G. GURZADYAN

1. Theory

Problem 1. The creation of a mathematical model of N-body gravitating system, enabling to either avoid compactness, measure and other difficulties or to define statistical properties (mixing, etc.) without use of those conditions. A more concrete aim could be formulated as follows:

Let $(X, \mathcal{B}(X), \mu, f)$ *be a smooth dynamical system with continuous or discrete time, where* X *is not compact and/or* $\mu(x) = \infty$.

The problem is to define:

a. the property of mixing;
b. correlation functions.

Problem 2. The study of behavior of time correlation functions for physical phase functions (kinetic energy, etc.). The strict formulation reads:

Consider a dynamical system when $\forall g_1, g_2 \in L^2(X) \quad \exists C_{g_1,g_2} > 0$ *and* $\beta_{g_1,g_2} > 0$ *such that* $\forall t > 0$ *one has*

$$| b_{g_1,g_2}(t) | \leq C_{g_1,g_2} \exp(-\beta_{g_1,g_2} t),$$

for $t \in R$ *or* $t \in Z$.

The problem then comes to:

a. finding out conditions which the dynamical system with such properties should fulfill;
b. an estimation of β.

Problem 3. The derivation of physical conditions to describe the core collapse, the evaporation and other evolutionary effects of N-body systems.

Problem 4. The study of the role of stochasticity and regularity of motion in determining the morphology of galaxies.

2. Computer Simulations

Problem 5. The building of a computer code to describe the N-body system with a phase trajectory close to that of the physical one for long enough time scales; the same for systems with non-point particles.

Let $x(t)$ be an exact solution of the N-body problem, $x_c(t)$ the one calculated by computer and by means of particular method, and

$$\varepsilon_c(T) = \sup_{t\in[0,T]} \|x_c(t) - x(t)\|.$$

The problem is to:

a. evaluate $\varepsilon_c(T)$;
b. study the limit

$$\varepsilon_c \equiv \lim_{t\to\infty} \sup \frac{\ln \varepsilon_c(T)}{T},$$

and its relation to Lyapunov characteristic exponents;
c. find out methods for which $\varepsilon_c = 0$.

Problem 6. The development of effective methods for numerical study of the statistical properties of N-body systems, particularly for the local (in time) characteristics of the instability.

Problem 7. The search of computer algebraic methods for the study of the evolution of gravitating systems, in order to avoid the numerical integration of differential equations (iterations).

3. Observations

Problem 8. The analysis of high accuracy data on central regions of galaxies and star clusters, in order to find out the dependence on radius of the number density of stars, velocity dispersion, the eccentricity of system, etc., and hidden relations between them.

Problem 9. The formulation of quantitative empirical relations determining the position of the stellar system on the path of evolution.

Problem 10. The search of empirical relations enabling to distinguish the role in relaxation driving effects of two-body and N-body gravitational interactions of stars in galaxies and star clusters.

Comments on "10 Key Problems"

Richard H. MILLER

1. General Remarks

Exponential separation of most (initially) neighboring trajectories in phase space is generally accepted in this audience, and coping with it is generally regarded as important in our understanding of stellar systems. However, we seem on several occasions to have become obsessed with the details at the expense of the broader picture.

The goal of studies in stellar dynamics is to understand the dynamics of real stellar systems—galaxies or star clusters. Logical connections in our understanding and interpretations of the dynamics of galaxies are shown in Fig. 1.

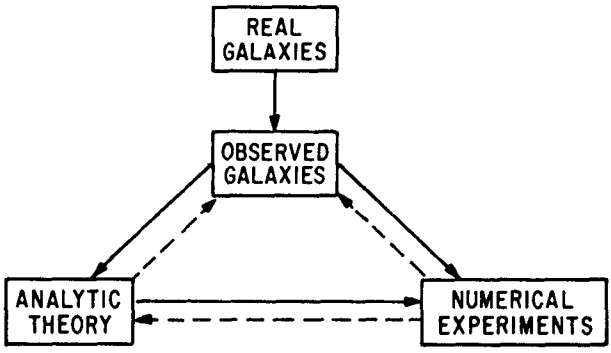


Fig. 1
Logical Connections

Real galaxies are the physical systems of interest; what little we know about them comes from observation, which rarely tells us about the physical properties in any direct way. Observations are interpreted both through analytic theory and numerical experiments, and we hope that observers, analytic theoreticians, and numerical experimenters talk to each other.

The crucial question for this conference is how trajectory separation affects our understanding of the physics of galaxies (or of star clusters). Does trajectory separation compromise our treatments by analytic theory? By numerical experiments? These questions do not seem to have been clarified at this workshop.

Numerical experiments are compromised only if the numerical trajectories differ in some important way from "physical trajectories". The difference is important only if numerical trajectories visit parts of the phase space differently from physical trajectories. That would imply some integrals present in one system but not in the other. It seems unlikely, but how do we address this problem?

A purely numerical approach, such as shadowing, won't help because all trajectories being compared in shadowing are computed trajectories. We can only say how well computed trajectories compare; but if computed trajectories visit different parts of the phase space differently from physical trajectories, neither the original nor the shadow trajectory can explore the phase space freely. Both are subject to the same restrictions.

2. Comment on Gurzadyan's Problem 2

Autocorrelations of some state functions are known *not* to decay exponentially with time in stellar dynamical systems. The question is not trivial, but the result has been known for some time. This problem was studied many years ago by Chandrasekhar as he was trying to develop a "statistical stellar dynamics." He envisioned the development of a stellar system as a kind of Brownian motion, in which the transition from the state of the system at time t to that at $t + \mathrm{d}t$ is a Markov process with some by a transition matrix. Markov processes lead to exponential decays in the autocorrelations of descriptive functions. Chandrasekhar arrived at the startling conclusion that the autocorrelation in the force acting at one point within a stellar system decayed as $1/t$, rather than as $\exp(-\beta t)$, as would be expected for a Markov process. The arguments are developed in a remarkable set of papers with von Neumann (Chandrasekhar & von Neumann 1942, 1943) followed by a couple by Chandrasekhar alone (Chandrasekhar 1944a, 1944b). The $1/t$ result is in (1944b). The conclusion was reconfirmed a few years ago by Ed Lee (Lee 1968), using more modern language.

Chandrasekhar's result is stronger than the demonstration that the gravitational N-body system is not an Anosov C-system that I mentioned earlier at this conference. My result says that the phase space does not have a simple geometric structure in which the number of expanding and contracting dimensions is constant throughout the space, which is quite a strong requirement. Chandrasekhar, on the other hand, demonstrated that the time dependence of the autocorrelation in the force acting at a given point is not exponential. His demonstration imposes less demanding requirements on the phase space.

3. Finite Numbers of Particles

The fact that numerical experiments handle a finite number of particles has been mentioned somewhat apologetically. One should not apologize. Real galaxies and real star clusters have a finite number of particles. A limited number of particles actually makes numerical studies a more faithful representation of real galaxies than are analytic Vlasov models with an infinite number of particles.

With modes, where we expect $\mathcal{O}(n)$ modes, numerical experiments may have half a million modes, while Vlasov models have an infinite number. Neither matches a real galaxy correctly, but we can hope that the low-order

modes will be similar between the two cases. This hope has been borne out in all cases where comparisons could be made, a result which gives confidence that both numerical experiments and analytic theory report those low-order modes correctly.

Chandrasekhar S., 1944a, ApJ 99, 25
Chandrasekhar S., 1944b, ApJ 99, 47
Chandrasekhar S., von Neumann J., 1942, ApJ 95, 489
Chandrasekhar S., von Neumann J., 1943, ApJ 97, 1
Lee, E. P., 1968, ApJ 151, 687

Avram HAYLI

My general feeling is that the very nature of the N-body problem will prevent from drawing precise previsions over the long term. Of course certain well behaved maps exhibit shadowing property, in the sense that near a numerically computed orbit in the phase space there exists a true orbit of different, yet unknown, initial condition than the one intended. But as we are looking for a precise description, I wonder if such a result may have any practical interest. Anyway the building of a computer code to describe the N-body system with a phase trajectory close to that of the theoretical one for arbitrarily long time scales will fail because of the repeated close encounters of three or more particles. Finding out a method to control ϵ_c on the long term and finally ask ϵ_c to be 0 seems in my opinion an impossible hope.

Louis MARTINET

It seems to me that my paper "On the Permissible Percentage of Chaotic Orbits in Various Morphological Types of Galaxies" brings some elements of answer to problem 4. However, a key approach towards a general solution could consist to translate the original behaviour into terms of behaviour of geodesics on a surface with a metric defined by means of the potential of the system considered. Practically nothing exists in the literature in this frame of mind concerning applications to galaxies. In fact the realistic galactic potentials are not easily tractable in this context. However, we plan to deal with some simple cases in the near future.

Yakov PESIN (Pennsylvania State University)

As for problems 1 and 2, I hope to get some results about the decay of correlations for general systems with non-zero Lyapunov exponents. There is also that one can say about infinite measure case. I am now much interested in a couple of things including fractal geometry and fractal dimension and spatial-temporal chaos. I have done something interesting and there is a great hope to get some strong results (but for the dissipative case so far) about

how one can obtain information on a dynamical system by looking at the lattice model one has while working with a computer. Another problem is to construct the gravitation theory (or something like that) on a fractal set (and I suppose that the Universe has a fractal structure). I have some tiny idea about that.

Juan Carlos MUZZIO

Concerning problems 5 and 6, on numerical codes for N-body systems, let us first distinguish two very different cases: a) Systems where the relaxation time is comparable to their age (e.g., open and globular clusters); b) Systems where the relaxation time is much larger than their age (e.g., galaxies). In the first case, the star-star interactions must be taken into account, so that each star in the system is represented with one body in the simulation and purely Newtonian forces are used; close encounters may demand the use of regularization techniques and, alternatively, the effect of distant masses can be averaged (as in tree codes). This problem is physically unstable: very small departures from a given initial condition will result in large departures from the final condition for the *real* system, if the evolution is followed for a long enough time; therefore, to build a stable code in this case is impossible, and we can only hope that *macroscopic* properties (say, half-mass radii, number of escapees, and so on) be preserved, even though *microscopic* properties (individual positions and velocities) are not. In case b), instead, each body in the simulation represents millions of stars in reality and, correspondingly, the individual masses and the interparticle distances are very different in the stellar system and in the simulation. Besides, in order to get reasonably small relaxation effects in the simulation, one has to resort to codes that compute the potential through expansion in suitable basis functions, or to direct summation codes with softened force laws and rather large softening parameters. It would be desirable to obtain in the simulation trajectories similar to those of stars in the equivalent general smooth constant potential but, again, that seems an impossible task: the very fact that the relaxation effects can be reduced, but not eliminated, shows that the energies of the individual particles are not conserved, as required by motion in such a potential. Thus, we can only hope for an accurate description of macroscopic properties in this case too. In the particular case of problem 6, I believe that the new perturbation particle methods offer great promise because, using particles only for the perturbation, they greatly reduce the relaxation effects that arise from the necessarily finite number of particles that enter in the simulations.

The increase in quantity and quality of observational data (problem 8) poses an interesting challenge to theoreticians. Most, perhaps all, of us agree nowadays with the idea, pioneered by K.R. Popper, that no scientific theory can be proved to be right, it can only be proved to be wrong: no matter how many observations corroborate the theory, if a new observation contradicts it, we must change the theory (or show that the new observation was wrong!).

Therefore, it is the duty of observers to find facts that could lead to the rejection of current theories, rather than to their corroboration, and no theoretician should feel bad about this. Alternatively, theoreticians should make predictions that could help the observers to disprove their theories and, in that way, aid themselves to build better theories. I think that it is not so important for a theory to be true (in the long run, all may be proved false), as it is for it to be fruitful: even if a theory turns out to be wrong, if in the process it suggested new observations that led to better theories, then it should be regarded as a very useful theory indeed. I believe that it is extremely important for theory and observation to march at a similar pace. If one advances much farther than the other, sooner or later the former will have to wait for the latter to come closer. One of my dearest Professors, the late "Don Miguel" Itzigsohn, used to remark that, while it is commonplace to note that Kepler would not have found his laws without the exquisitely precise (for their time) observations of Tycho Brahe, it is equally true that, had Tycho's observations been even more precise, Kepler could not have found his laws either, because it would have then been obvious that the planetary orbits are not exactly ellipses. The new observational material is already here, waiting to be used to devise more refined theories: the time is now, the theoreticians are us; otherwise, future technical advances will yield even better and more abundant observations that will be ever more difficult to accommodate within the present theoretical framework.

Shogo INAGAKI

It is sometimes too difficult to study the chaotic properties of real N-body systems because the force diverges at zero distance and the phase space extends to infinity. Therefore I would suggest to study simpler models such as I suggested in the workshop before studying real N-body systems. Though my model has some similar properties as real N-body systems, it has finite forces at all distances and the configuration space is compact and one-dimensional. Therefore it should be much easier to deal with.

It is also important to clarify the meaning of Gurzadyan-Savvidy time-scale.

Daniel PFENNIGER

My comments are related to Problem 4.

The problem that should be examined in any theories is to determine its fragile points. Often this task may take a long time, because a new theory is first tested against the simplest cases. Next more subtle aspects can be discovered and investigated.

Concerning the N-body model with respect to stellar systems, in the last decades one "subtle" point has become clear: chaos makes models fragile to

perturbations. As consequence such questions should be asked: what is the scope of applicability of Liouville's theorem in chaotic stellar systems, when each trajectory or the global system is sensitive to perturbations from the rest of the Universe?

In an earlier work with Colin Norman (1990, ApJ 363, 391) I was surprised to see how chaotic orbits are also responsive to dissipative perturbations. Weak dissipative effects are *amplified* by chaos. Most stellar systems do have a weak degree of dissipation and are strongly chaotic. Could a weak dissipation determines the long term state of stellar systems? In gas rich systems like spiral galaxies this is likely, but ellipticals contain also several percent of gas. Other secular dissipative effects are related to the mass loss from stars; clearly the mass lost in a stellar population after a few Gyr is not negligible.

Therefore, I doubt that "phase space volume conservation" arguments can really be applied to processes like galaxy formation by collapses or mergers of galaxies.

George S. DJORGOVSKI

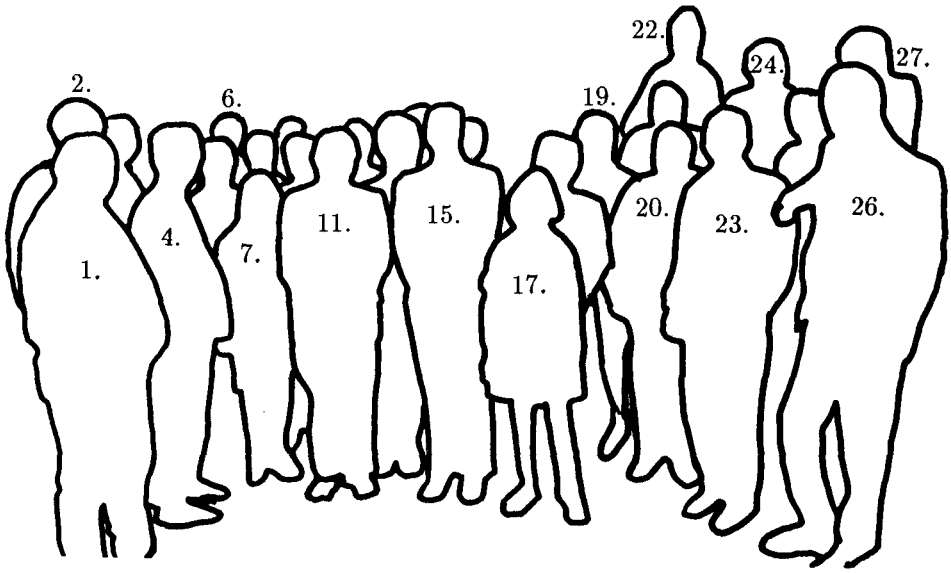
I wonder if we are ready for the 10 key questions, be it these or some others. This meeting has been a success – if for no other reason, then as a very interesting experiment in the sociology of science – but we still have a long way to go! Among the mathematicians, numerical simulators, and observers, we barely even have a common language, or maybe even the common goals. Still, we have seen lots of mutual good will to learn from each other, and to understand each other. I guess we need more meetings to do our compulsory relaxation and strong mixing, and I don't mean only the drinks.

Let us cast the scene as a three-body interaction between the mathematical ergodic theory, numerical simulations (disturbingly few of which at this conference actually dealt with modeling of stellar systems!), and observations. These three things are not isomorphic, and the best they can hope is to make the life more interesting for each other, and maybe find some real-world manifestations which can be understood or plausibly explained by the theory. There are possible pitfalls all around. Mathematicians can get enamored by cleverness for its own sake, which does little good to anyone else, and it sure will not bring them any appreciation from the philistine masses of observers and suchlike lumpenproletariat. Similarly, numerical simulators can easily get lost in their fancy video games, and I have seen many tools in search of the problems, and most problems don't fit the tools. Finally, it is all too common for observers to wallow in a gross empiricism, or to pursue botanical astronomy as a hobby.

Well, speaking from my cultural bias, I think that there are some excellent problems out there, which could profit from the talents and expertise of black-belt dynamicists. Observations of real-life stellar systems pose some fascinating challenges, and I tried to point out a few of them in my written contribution to this volume. I'd rather see a real problem tackled, no matter

how fuzzy, no matter how simply, than an unrealistic (and I'd say, usually sterile) toy problem. The real universe is far more interesting to some of us, than any logical construct involving grids of perfect balls and springs or similar contraptions operating without a benefit of friction, external perturbations, and similar annoyances which only exist in the real world... I think that there is a real payoff for an adventuresome dynamicist who manages to solve a problem posed by the real world, both in terms of an intellectual satisfaction, and a professional recognition.

But remember: there is no such thing as an isolated or a dissipationless stellar system out there. The universe is a chaotic mess, maybe even mathematically so.



Half the participants sacrificed one minute coffee-break for the photo! From left to right:

1. Evgenij NEZHINSKIJ	2. Włodzimierz GODŁOWSKY	3. Tetsuro KONISHI
4. Toshio TSUCHIYA	5. Naoteru GOUDA	6. David EARN
7. Yoko FUNATO	8. Gerald QUINLAN	9. Haywood SMITH ?
10. Hervé WOZNIAK	11. Tjeerd VAN ALBADA	12. Jean-Louis ROUET
13. Junichiro MAKINO	14. Roger FUX	15. Vahe GURZADYAN
16. Leonid OSSIPKOV	17. Diana GURZADYAN	18. Nelson TABIRYAN
19. Juan MUZZIO	20. Jean HEYVAERTS	21. Richard MILLER
22. Marc BALCELLS	23. Geo MEYLAN	24. Jacques LASKAR
25. Lia ATHANASSOULA	26. Daniel PFENNIGER	27. Shogo INAGAKI

Indexes

Author Index

Subject Index

Lecture Notes in Physics

For information about Vols. 1–390
please contact your bookseller or Springer-Verlag

Vol. 391: A. Zdziarski, M. Sikora (Eds.), Ralativistic Hadrons in Cosmic Compact Objects. Proceedings, 1990. XII, 182 pages. 1991.

Vol. 392: J.-D. Fournier, P.-L. Sulem (Eds.), Large-Scale Structures in Nonlinear Physics. Proceedings. VIII, 353 pages. 1991.

Vol. 393: M. Remoissenet, M.Peyrard (Eds.), Nonlinear Coherent Structures in Physics and Biology. Proceedings. XII, 398 pages. 1991.

Vol. 394: M. R. J. Hoch, R. H. Lemmer (Eds.), Low Temperature Physics. Proceedings. X,374 pages. 1991.

Vol. 395: H. E. Trease, M. J. Fritts, W. P. Crowley (Eds.), Advances in the Free-Lagrange Method. Proceedings, 1990. XI, 327 pages. 1991.

Vol. 396: H. Mitter, H. Gausterer (Eds.), Recent Aspects of Quantum Fields. Proceedings. XIII, 332 pages. 1991.

Vol. 398: T. M. M. Verheggen (Ed.), Numerical Methods for the Simulation of Multi-Phase and Complex Flow. Proceedings, 1990. VI, 153 pages. 1992.

Vol. 399: Z. Švestka, B. V. Jackson, M. E. Machedo (Eds.), Eruptive Solar Flares. Proceedings, 1991. XIV, 409 pages. 1992.

Vol. 400: M. Dienes, M. Month, S. Turner (Eds.), Frontiers of Particle Beams: Intensity Limitations. Proceedings, 1990. IX, 610 pages. 1992.

Vol. 401: U. Heber, C. S. Jeffery (Eds.), The Atmospheres of Early-Type Stars. Proceedings, 1991. XIX, 450 pages. 1992.

Vol. 402: L. Boi, D. Flament, J.-M. Salanskis (Eds.), 1830-1930: A Century of Geometry. VIII, 304 pages. 1992.

Vol. 403: E. Balslev (Ed.), Schrödinger Operators. Proceedings, 1991. VIII, 264 pages. 1992.

Vol. 404: R. Schmidt, H. O. Lutz, R. Dreizler (Eds.), Nuclear Physics Concepts in the Study of Atomic Cluster Physics. Proceedings, 1991. XVIII, 363 pages. 1992.

Vol. 405: W. Hollik, R. Rückl, J. Wess (Eds.), Phenomenological Aspects of Supersymmetry. VII, 329 pages. 1992.

Vol. 406: R. Kayser, T. Schramm, L. Nieser (Eds.), Gravitational Lenses. Proceedings, 1991. XXII, 399 pages. 1992.

Vol. 407: P. L. Smith, W. L. Wiese (Eds.), Atomic and Molecular Data for Space Astronomy. VII, 158 pages. 1992.

Vol. 408: V. J. Martínez, M. Portilla, D. Sàez (Eds.), New Insights into the Universe. Proceedings, 1991. XI, 298 pages. 1992.

Vol. 409: H. Gausterer, C. B. Lang (Eds.), Computational Methods in Field Theory. Proceedings, 1992. XII, 274 pages. 1992.

Vol. 410: J. Ehlers, G. Schäfer (Eds.), Relativistic Gravity Research. Proceedings, VIII, 409 pages. 1992.

Vol. 411: W. Dieter Heiss (Ed.), Chaos and Quantum Chaos. Proceedings, XIV, 330 pages. 1992.

Vol. 412: A. W. Clegg, G. E. Nedoluha (Eds.), Astrophysical Masers. Proceedings, 1992. XX, 480 pages. 1993.

Vol. 413: Aa. Sandqvist, T. P. Ray (Eds.); Central Activity in Galaxies. From Observational Data to Astrophysical Diagnostics. XIII, 235 pages. 1993.

Vol. 414: M. Napolitano, F. Sabetta (Eds.), Thirteenth International Conference on Numerical Methods in Fluid Dynamics. Proceedings, 1992. XIV, 541 pages. 1993.

Vol. 415: L. Garrido (Ed.), Complex Fluids. Proceedings, 1992. XIII, 413 pages. 1993.

Vol. 416: B. Baschek, G. Klare, J. Lequeux (Eds.), New Aspects of Magellanic Cloud Research. Proceedings, 1992. XIII, 494 pages. 1993.

Vol. 417: K. Goeke P. Kroll, H.-R. Petry (Eds.), Quark Cluster Dynamics. Proceedings, 1992. XI, 297 pages. 1993.

Vol. 418: J. van Paradijs, H. M. Maitzen (Eds.), Galactic High-Energy Astrophysics. XIII, 293 pages. 1993.

Vol. 419: K. H. Ploog, L. Tapfer (Eds.), Physics and Technology of Semiconductor Quantum Devices. Proceedings, 1992. VIII, 212 pages. 1993.

Vol. 420: F. Ehlotzky (Ed.), Fundamentals of Quantum Optics III. Proceedings, 1993. XII, 346 pages. 1993.

Vol. 421: H.-J. Röser, K. Meisenheimer (Eds.), Jets in Extragalactic Radio Sources. XX, 301 pages. 1993.

Vol. 422: L. Päivärinta, E. Somersalo (Eds.), Inverse Problems in Mathematical Physics. Proceedings, 1992. XVIII, 256 pages. 1993.

Vol. 423: F. J. Chinea, L. M. González-Romero (Eds.), Rotating Objects and Relativistic Physics. Proceedings, 1992. XII, 304 pages. 1993.

Vol. 424: G. F. Helminck (Ed.), Geometric and Quantum Aspects of Integrable Systems. Proceedings, 1992. IX, 224 pages. 1993.

Vol. 425: M. Dienes, M. Month, B. Strasser, S. Turner (Eds.), Frontiers of Particle Beams: Factories with e^+ e^- Rings. Proceedings, 1992. IX, 414 pages. 1994.

Vol. 426: L. Mathelitsch, W. Plessas (Eds.), Substructures of Matter as Revealed with Electroweak Probes. Proceedings, 1993. XIV, 441 pages. 1994

Vol. 427: H. V. von Geramb (Ed.), Quantum Inversion Theory and Applications. Proceedings, 1993. VIII, 481 pages. 1994.

Vol. 428: U. G. Jørgensen (Ed.), Molecules in the Stellar Environment. Proceedings, 1993. VIII, 440 pages. 1994.

Vol. 429: J. L. Sanz, E. Martínez-González, L. Cayón (Eds.), Present and Future of the Cosmic Microwave Background. Proceedings, 1993. VIII, 233 pages. 1994.

Vol. 430: V. G. Gurzadyan, D. Pfenniger (Eds.), Ergodic Concepts in Stellar Dynamics. Proceedings, 1993. XVI, 302 pages. 1994.

New Series m: Monographs

Vol. m 1: H. Hora, Plasmas at High Temperature and Density. VIII, 442 pages. 1991.

Vol. m 2: P. Busch, P. J. Lahti, P. Mittelstaedt, The Quantum Theory of Measurement. XIII, 165 pages. 1991.

Vol. m 3: A. Heck, J. M. Perdang (Eds.), Applying Fractals in Astronomy. IX, 210 pages. 1991.

Vol. m 4: R. K. Zeytounian, Mécanique des fluides fondamentale. XV, 615 pages, 1991.

Vol. m 5: R. K. Zeytounian, Meteorological Fluid Dynamics. XI, 346 pages. 1991.

Vol. m 6: N. M. J. Woodhouse, Special Relativity. VIII, 86 pages. 1992.

Vol. m 7: G. Morandi, The Role of Topology in Classical and Quantum Physics. XIII, 239 pages. 1992.

Vol. m 8: D. Funaro, Polynomial Approximation of Differential Equations. X, 305 pages. 1992.

Vol. m 9: M. Namiki, Stochastic Quantization. X, 217 pages. 1992.

Vol. m 10: J. Hoppe, Lectures on Integrable Systems. VII, 111 pages. 1992.

Vol. m 11: A. D. Yaghjian, Relativistic Dynamics of a Charged Sphere. XII, 115 pages. 1992.

Vol. m 12: G. Esposito, Quantum Gravity, Quantum Cosmology and Lorentzian Geometries. Second Corrected and Enlarged Edition. XVIII, 349 pages. 1994.

Vol. m 13: M. Klein, A. Knauf, Classical Planar Scattering by Coulombic Potentials. V, 142 pages. 1992.

Vol. m 14: A. Lerda, Anyons. XI, 138 pages. 1992.

Vol. m 15: N. Peters, B. Rogg (Eds.), Reduced Kinetic Mechanisms for Applications in Combustion Systems. X, 360 pages. 1993.

Vol. m 16: P. Christe, M. Henkel, Introduction to Conformal Invariance and Its Applications to Critical Phenomena. XV, 260 pages. 1993.

Vol. m 17: M. Schoen, Computer Simulation of Condensed Phases in Complex Geometries. X, 136 pages. 1993.

Vol. m 18: H. Carmichael, An Open Systems Approach to Quantum Optics. X, 179 pages. 1993.

Vol. m 19: S. D. Bogan, M. K. Hinders, Interface Effects in Elastic Wave Scattering. XII, 182 pages. 1994.

Vol. m 20: E. Abdalla, M. C. B. Abdalla, D. Dalmazi, A. Zadra, 2D-Gravity in Non-Critical Strings. IX, 319 pages. 1994.