UNDERSTANDING COMPLEX SYSTEMS

Springer: COMPLEXITY

Marco Thiel
Jürgen Kurths
M. Carmen Romano
Alessandro Moura
György Károlyi
Editors

Nonlinear Dynamics and Chaos: Advances and Perspectives

Springer

Springer Complexity

Springer Complexity is an interdisciplinary program publishing the best research and academic-level teaching on both fundamental and applied aspects of complex systems – cutting across all traditional disciplines of the natural and life sciences, engineering, economics, medicine, neuroscience, social and computer science.

Complex Systems are systems that comprise many interacting parts with the ability to generate a new quality of macroscopic collective behavior the manifestations of which are the spontaneous formation of distinctive temporal, spatial or functional structures. Models of such systems can be successfully mapped onto quite diverse "real-life" situations like the climate, the coherent emission of light from lasers, chemical reaction-diffusion systems, biological cellular networks, the dynamics of stock markets and of the internet, earthquake statistics and prediction, freeway traffic, the human brain, or the formation of opinions in social systems, to name just some of the popular applications.

Although their scope and methodologies overlap somewhat, one can distinguish the following main concepts and tools: self-organization, nonlinear dynamics, synergetics, turbulence, dynamical systems, catastrophes, instabilities, stochastic processes, chaos, graphs and networks, cellular automata, adaptive systems, genetic algorithms and computational intelligence.

The two major book publication platforms of the Springer Complexity program are the monograph series "Understanding Complex Systems" focusing on the various applications of complexity, and the "Springer Series in Synergetics", which is devoted to the quantitative theoretical and methodological foundations. In addition to the books in these two core series, the program also incorporates individual titles ranging from textbooks to major reference works.

Understanding Complex Systems

Founding Editor: J.A. Scott Kelso

Future scientific and technological developments in many fields will necessarily depend upon coming to grips with complex systems. Such systems are complex in both their composition – typically many different kinds of components interacting simultaneously and nonlinearly with each other and their environments on multiple levels – and in the rich diversity of behavior of which they are capable.

The Springer Series in Understanding Complex Systems series (UCS) promotes new strategies and paradigms for understanding and realizing applications of complex systems research in a wide variety of fields and endeavors. UCS is explicitly transdisciplinary. It has three main goals: First, to elaborate the concepts, methods and tools of complex systems at all levels of description and in all scientific fields, especially newly emerging areas within the life, social, behavioral, economic, neuro- and cognitive sciences (and derivatives thereof); second, to encourage novel applications of these ideas in various fields of engineering and computation such as robotics, nano-technology and informatics; third, to provide a single forum within which commonalities and differences in the workings of complex systems may be discerned, hence leading to deeper insight and understanding.

UCS will publish monographs, lecture notes and selected edited contributions aimed at communicating new findings to a large multidisciplinary audience.

Marco Thiel · Jürgen Kurths ·
M. Carmen Romano · Alessandro Moura ·
György Károlyi
Editors

Nonlinear Dynamics and Chaos: Advances and Perspectives

Springer

Editors
Marco Thiel
Institute of Complex Systems and
Mathematical Biology
University of Aberdeen
King's College, Meston Building
AB24 3UE Aberdeen
United Kingdom
m.thiel@abdn.ac.uk

M. Carmen Romano
Institute of Complex Systems and
Mathematical Biology
University of Aberdeen
King's College, Meston Building
AB24 3UE Aberdeen
United Kingdom
m.roman@abdn.ac.uk

György Károlyi
UCB Magyarorszag KFT
Hüvösvölgyi ut 54
1021 Budapest
Hungary
karolyi@tas.me.bme.hu

Jürgen Kurths
Potsdam Institute for Climate Impact
Research (PIK)
Telegrafenberg A31
14473 Potsdam
Juergen.Kurths@pik-potsdam.de

Alessandro Moura
Institute of Complex Systems and
Mathematical Biology
University of Aberdeen
King's College, Meston Building
AB24 3UE Aberdeen
United Kingdom
a.moura@abdn.ac.uk

ISSN 1860-0832 e-ISSN 1860-0840
ISBN 978-3-642-04628-5 e-ISBN 978-3-642-04629-2
DOI 10.1007/978-3-642-04629-2
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2010925336

Cover design: WMXDesign, Heidelberg

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

This book is dedicated to Celso Grebogi,
who has been a constant source of inspiration
and encouragement to those who know him.

Preface

This book is a collection of papers contributed by some of the greatest names in the areas of chaos and nonlinear dynamics. Each paper examines a research topic at the frontier of the area of dynamical systems. As well as reviewing recent results, each paper also discusses the future perspectives of each topic. The result is an invaluable snapshot of the state of the field by some of the most important researchers in the area.

The first contribution in this book (the section entitled "How did you get into Chaos?") is actually not a paper, but a collection of personal accounts by a number of participants of the conference held in Aberdeen in September 2007 to honour Celso Grebogi's 60th birthday. At the instigation of James Yorke, many of the most well-known scientists in the area agreed to share their tales on how they got involved in chaos during a celebratory dinner in Celso's honour during the conference. This was recorded in video, we felt that these accounts were a valuable historic document for the field. So we decided to transcribe it and include it here as the first section of the book.

The dynamics of maps on the complex plane provide some of the most striking examples of chaos and fractal invariant sets in dynamics, and has been of great importance to the field because they are amenable to rigorous treatment. The first paper in the book is R. Devaney's investigation of the dynamical properties of singularly-perturbed complex maps. He investigates Julia sets and other related sets which arise in maps with a pole, and classifies their dynamics.

One of the most exciting developments in recent years is the application of dynamical systems techniques to complex networks of interacting components, each having their own internal dynamics, and each being coupled to other nodes. P. Ashwin, G. Orosz and J. Borresen review how complex dynamics can arise even in simple, fully symmetric and globally coupled networks. They make the important point that not only the network topology (which is usually emphasised in the literature), but also the properties of the coupling function are crucial to determine the system's global dynamics.

Fluid dynamics is an area that has always had a close relation with chaos. The motion of particles advected by time-dependent flows is a prime example of a chaotic system, and chaotic advection has been observed in many beautiful experiments. Most of the existing theoretical work considers advected particles as having

vanishing size, even though it is known that their finite size can have considerable consequences for their dynamics. J. Cartwright *et al.* review the dynamics of finite-size particles in chaotic flows. This is the first published review of this important subject.

Hamiltonian dynamics occupies a special place in the area of dynamical systems, because of its applications to classical mechanics, celestial mechanics, physics and other areas. In many Hamiltonian systems, there is a clear separation of slow and fast degrees of freedom, and it is common practice to model the effects of the fast variables by noise and damping, which results in a Langevin equation for the slow degrees of freedom. However, the rigorous mathematical foundations for this are not well-established. R. MacKay proposes a rigorous way to derive a Langevin equation for Hamiltonian systems, by assuming that the fast variables have an Anosov mixing dynamics.

In their contribution, A. Politi and A. Torcini review the concepts of *stable chaos*, that is, the presence of irregular behaviour even though the dynamics is still locally stable. Although the irregular behaviour is transient in these cases, the transient's lifetime diverges in the thermodynamic limit, which makes stable chaos relevant for out-of-equilibrium statistical mechanics. The authors emphasise the connection between stable chaos and the spatio-temporal chaos shown by a class of cellular automata.

Chaotic transients are also the topic of the next paper, by Y.-C. Lai. He reviews the subject of superpersistent chaotic transients, which refers to the extremely long lifetimes of some systems near bifurcations. He explains the dynamical origins of these long lifetimes, and how they are related to the riddling bifurcation and riddled basins. He also discusses the occurrence of transients in spatially extended systems, and the application of these concepts to the motion of particles in fluid flows.

One of the biggest current topics of research in the field of dynamical systems is synchronisation, and the next three papers are all concerned with this fascinating subject. The paper by P. Read and A. Castrajón-Pita investigates the possibility of synchronisation in spatio-temporal systems, focusing on the Earth's climate system. They make a good case for the possibility of the existence of synchronised oscillations in weather patterns. Experiments are performed in fluid-dynamical systems which are analogues of major components of our planet's weather system, and synchronisation is indeed observed.

Many of the systems in which the concept of synchronisation is important are noisy, especially biological systems at the cellular level. It is therefore very important to understand how synchronisation works in the presence of noise. R. Ramaswami *et al.* give an overview of stochastic synchronisation, giving a number of examples including chemical reactions and gene regulation networks.

Synchronisation was first discovered due to the observation by Christian Huygens that two clocks connected to the same wooden beam eventually synchronise their oscillations. A. Pogromski, D. Rijlaarsdam and H. Nijmeijer present an experimental setup which allows a thorough exploration of the synchronisation of mechanical systems, in the spirit of Huygens's original observation but in a much more controlled way. Their setup allows the coupling between two mechanical oscillators

to be controlled, and thus makes it possible to study different kinds of synchronisation phenomena, such as the synchronisation of pendula, rotating objects, etc. They describe results of synchronisation experiments with two Duffing oscillators and with two rotating disks.

The control of chaos is one of many areas to which Celso Grebogi has fundamental contributions. G. Riga, S. Lenci and J. M. T. Thompson's paper reviews the OGY theory of controlling chaos from a historical perspective, and discusses applications in mechanics and related fields. They compare and contrast the OGY strategy based on stabilising a single periodic orbit of the chaotic set with a global method of trying to stabilise the overall system dynamics.

Since the seminal embedding theorems by Takens and others, time-series analysis has been an important area within dynamical systems. Closing the book is R. Stoop and M. Christen's investigation on methods to extract regular patterns from time-series with noisy background. This is a crucial problem in neuroscience and other fields, and one in which traditional methods, such as power-spectrum and related procedures, usually fail. They propose a method based on staircase-like structures in the correlation plot, and derive a number of analytical results on their method, which suggest that their method is very promising in applications.

The contributions in this book cover a broad range of topics within the large area of dynamical systems and chaos, and they range from pure mathematics to real-world applications. They show that our field is as exciting as ever, and has a brilliant future ahead of it.

In September 2007, a conference was held in Aberdeen, Scotland, to celebrate the 60th birthday of Celso Grebogi. The list of invited speakers, among them many of the contributors to this book, reads like a "who's who" of the areas of dynamical systems and chaos. This conference was the first time so many great names in the field have been gathered together in a single event in decades, and it presented a unique opportunity to assess the present state of the area and its future directions. It was felt by many of the participants of that event that a book with in-depth surveys of important topics in the field was timely, and this was the driving force for putting together this volume.

Aberdeen, UK
April 2010

György Károlyi, Jürgen Kurths, Alessandro Moura, Marco Thiel, M. Carmen Romano

Contents

Contributors

Peter Ashwin College of Engineering, Mathematics and Physical Sciences, Mathematics Research Institute, University of Exeter, Exeter, EX4 4QF, UK, P.Ashwin@exeter.ac.uk

Jon Borresen Department of Computing and Mathematics, Manchester Metropolitan University, Manchester, M1 5GD, UK, J.Borresen@mmu.ac.uk

Julyan H. E. Cartwright Instituto Andaluz de Ciencias de la, Tierra, CSIC-Universidad de Granada, E-18071 Granada, Spain, julyan@lec.csic.es

Alfonso A. Castrejón-Pita Atmospheric, Oceanic & Planetary Physics, Clarendon Laboratory, Department, of Physics, University of Oxford, Oxford, OX1 3PU, UK, aacp@atm.ox.ac.uk

Markus Christen Institute of Neuroinformatics, Swiss Federal Institute of Technology, of Zurich ETHZ, University of Zurich, CH-8057 Zurich, Switzerland, markus.christen@pantaris.ch

Robert L. Devaney Department of Mathematics, Boston University, Boston MA 02215, USA, bob@bu.edu

Ulrike Feudel Institut für Chemie und Biologie des, Meeres, Carl von Ossietzky Universität, Oldenburg, PF 2503, D-26111 Oldenburg, Germany, ulrike.feudel@uni-oldenburg.de

György Károlyi Department of Structural Mechanics, Budapest University of Technology, and Economics, H-1111 Budapest, Hungary; Centre for Applied Dynamics Research, School of Engineering, King's College, University of Aberdeen, AB24 3UE, Scotland, UK, karolyi@tas.me.bme.hu

Jürgen Kurths Potsdam Institute for Climate, Impact Research, 14473 Potsdam, Germany, kurths@pik-potsdam.de

Ying-Cheng Lai School of Electrical, Computer and Energy Engineering, Department of Physics, Arizona State University, Tempe, Arizona 85287, USA; Institute for Complex Systems and Mathematical Biology, School of

Natural and Computing Sciences, King's College, University of Aberdeen, UK, Ying-Cheng.Lai@asu.edu

Stefano Lenci Dipartimento di Architettura, Costruzioni e Strutture, Universit'a Politecnica delle Marche, I-60131 Ancona, Italy, lenci@univpm.it

Robert S. MacKay Mathematics Institute and Centre for Complexity Science, University of Warwick, Coventry CV4 7AL, UK, R.S.MacKay@warwick.ac.uk

Alessandro de Moura Institute of Complex Systems and Mathematical Biology, University of Aberdeen, King's College, Meston Building, AB24 3UE Aberdeen, United Kingdom, a.moura@abdn.ac.uk

Henk Nijmeijer Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands, h.nijneijer@tue.nl

Gábor Orosz School of Engineering, Computing and Mathematics, Mathematics Research Institute, University of Exeter, Exeter, EX4 4QF, UK; Department of Mechanical Engineering, University of California, Santa Barbara, California, 93106, USA, gabor@engineering.ucsb.edu

Oreste Piro Instituto de Física Interdisciplinar y, Sistemas Complejos, Universitat de les Illes Balears, E-07122 Palma de Mallorca, Spain, piro@ifisc.uib.es

Alexander Pogromsky Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands, A.Pogromsky@tue.nl

Antonio Politi Istituto dei Sistemi Complessi, Consiglio Nazionale delle Ricerche, I-50019 Sesto Fiorentino, Italy, antonio.politi@isc.cnr.it

Ram Ramaswamy Center for Computational Biology, and Bioinformatics, and School of Physical Sciences, Jawaharlal Nehru University, New Delhi 110067, India, r.ramaswamy@mail.jnu.ac.in

Peter L. Read Atmospheric, Oceanic & Planetary Physics, Clarendon Laboratory, Department, of Physics, University of Oxford, Oxford, OX1 3PU, UK, p.read1@physics.ox.ac.uk

Giuseppe Rega Dipartimento di Ingegneria Strutturale, e Geotecnica, Sapienza Universitá di Roma, I-00197 Rome, Italy, giuseppe.rega@uniroma1.it

David Rijlaarsdam Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands, D.J.Rijlaarsdam@tue.nl

M. Carmen Romano Institute of Complex Systems and Mathematical Biology, University of Aberdeen, King's College, Meston Building, AB24 3UE Aberdeen, United Kingdom, m.romano@abdn.ac.uk

R. K. Brojen Singh Centre for Interdisciplinary Research, in Basic Sciences, Jamia Milia Islamia, New Delhi 110025, India, rk.brojen@gmail.com

Ruedi Stoop Institute of Neuroinformatics, Swiss Federal Institute of Technology, of Zurich ETHZ, University of Zurich, CH-8057 Zurich, Switzerland, ruedi@ini.phys.ethz.ch

Tamás Tél Institute for Theoretical Physics, Eötvös University, H-1518 Budapest, Hungary, tel@general.elte.hu

Marco Thiel Institute of Complex Systems and Mathematical Biology, University of Aberdeen, King's College, Meston Building, AB24 3UE Aberdeen, United Kingdom, m.thiel@abdn.ac.uk

J. M. T. Thompson Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, Cambridge, CB3 0WA, UK, jmtt@ucl.ac.uk

Alessandro Torcini Istituto dei Sistemi Complessi, Consiglio Nazionale delle Ricerche; Centro Studi Dinamiche Complesse; Istituto Nazionale Fisica Nucleare – Sezione di Firenze, I-50019 Sesto Fiorentino, Italy, alessandro.torcini@isc.cnr.it

Changsong Zhou Department of Physics, Centre for Nonlinear Studies, The Beijing-Hong Kong-Singapore Joint Centre for Nonlinear and Complex Systems (Hong Kong), Hong Kong Baptist University, Kowloon Tong, Hong Kong, China, cszhou@hkbu.edu.hk

How Did You Get into Chaos?

Abstract In the celebratory dinner honouring Celso Grebogi's 60th birthday, a number of scientists in the area of chaos were asked by James Yorke to tell the tale about how they got involved in the field. Since all the participants have played crucial roles in the development of the subject, their stories give unique insights into the historical development of dynamical systems and chaos. We have transcribed their tales here.

James Yorke (Institute for Physical Sciences and Technology, University of Maryland, USA) – I would like to talk about the history of chaos, or my history of chaos. And to me, it seems like I have always known about chaos in the following sense. When I was in high school, I read several books by Norbert Wiener. Norbert Wiener wrote very technical books, which I could not read, but he also wrote non-technical books, in which he covered topics like the gas laws. So when I was in high school, I tried to show – understand this is "tried", I did not seem to do anything – I tried to show that maximal distribution was invariant under collisions. And so I didn't do anything on that but it prepared me for chaos. So that's my story. Where's Harry? Come on up, Harry. Harry is going to tell his story. I guess, some people have volunteered, twisting their arms a bit, but I am going to ask as many people who want to comment briefly on how they got into chaos, what was the motivation. And Harry is number one – Harry.

Harry Swinney (Department of Physics, University of Texas at Austin, USA) – In 1974, Jerry Gollub was at Haverford College and I was at the City College of New York. Jerry had a fellowship to go on leave, and said he would like to come up to City College, where we could do an experiment together. We talked about the kinds of experiments we might do, and finally decided to study a transition from one state to another in a non-equilibrium system. We chose the flow between concentric cylinders with the inner one rotating [the Couette-Taylor system] because we had read an interesting 1965 *Journal of Fluid Mechanics* paper by Don Coles in which he had studied the transition to time-dependent flow from the time-independent Taylor vortices that form in this system.

We were interested in understanding transitions that would occur with increasing Reynolds number. We found a 1943 paper by Lev Landau about sequences of

M. Thiel et al. (eds.), *Nonlinear Dynamics and Chaos: Advances and Perspectives,*
Understanding Complex Systems, DOI 10.1007/978-3-642-04629-2_1,

transitions in fluids. Landau said that as you increased the Reynolds number of a fluid, one would generically go from a time-independent state to one that was time-dependent with a single characteristic frequency describing its motion. With further increase in the Reynolds number, there would be a second frequency that would appear at a well-defined Reynolds number, then another transition with a third characteristic frequency, etc. Each new transition would occur at a well-defined Reynolds number and would bring in a frequency that was incommensurate with the previous frequencies. After a number of such transitions, one would have a flow with many characteristic incommensurate frequencies. That would be a very complicated flow, and that's what Landau called turbulence.

Coles' paper described the transition from time-independent Taylor vortices to vortices that have waves travelling around the cylinder; that is, the flow had a single characteristic frequency. Coles went on to say that as you increase the rotation speed of the cylinder, the flow becomes more complex, apparently in the way that Landau had predicted. Gollub and I scoured the literature to see if there was more evidence for a sequence of transitions to turbulent flow characterised by many frequencies, and we couldn't find any papers where the frequencies were actually measured. [James Yorke- You have two more minutes. Laughing in audience.] I'm just getting started. [Laughter]

I knew how to measure frequencies because I had the good fortune in 1960–1962 to work for the National Security Agency, where I learned about Fourier transforms and power spectra, because NSA was interested in analysing signals from certain places. I learned about the Nyquist theorem and how to get a good signal to noise ratio.

Gollub and I needed to compute Fourier transforms, but they were costly to compute with the limited computers we had. I had a friend who worked at Bell Labs who told me about an algorithm called the fast Fourier transform that had been developed by Cooley and Tukey. He got me deck of punched IBM cards with the Cooley – Tukey algorithm, which Gollub and I used to compute Fourier transforms of our data. We found a single fundamental frequency for the wavy vortex flow studied by Coles. As we cranked up the speed of the inner cylinder, we saw at a higher, well-defined Reynolds number a second frequency appeared in the power spectrum. This was just as we expected from Landau's paper. Then we cranked up the speed of the inner cylinder further and looked for a third frequency to come in, but instead we found the spectrum became noisy. We published a paper on the transition to turbulence in *Physical Review Letters*, saying that there appeared one frequency and then a second frequency, and then just noise, no more discrete frequencies.

Joel Lebowitz at Yeshiva University, not far from the City College of New York, heard about our work and called me and said "I have a visitor here at Yeshiva who might be interested in that experiment." So he brought this French fellow to our lab; his name was David Ruelle. Audience–Belgian. Harry Swinney- a Belgian, working in France. And Ruelle said: "You have seen evidence of behaviour described by a strange attractor." We didn't know what a strange attractor was, but that sounded good.

Then we heard of a professor named James Yorke at the University of Maryland, who was studying non-periodic behaviour in deterministic systems, and I called him

up. [Jim, do you remember?] [James Yorke – No.] [Laughter] Jerry Gollub and I got on a train to Maryland; this was September 1975. I never had gone to talk shop with a real mathematician before. [Laughter] We presented our data to Professor Yorke, and he said, "That's chaos." [Laughter, applause].

James Yorke – So my second volunteer, if you walk over here.

Floris Takens (Department of Mathematics, University of Groningen, The Netherlands) – Thank you. I was ordered yesterday to be a volunteer here and to say something about how I was involved in this chaos business. Well, I come from the other side, from mathematics. I had just finished my thesis in differential topology and decided to move to dynamical systems. This subject at that time was hardly existing in the Netherlands so I decided to spend a year at IHES, Bures-sur-Ytette in France. There were some foreign visitors there, like Steve Smale, Mike Shub, Charles Pugh, and others. I started to learn dynamical systems there. There were seminars and a course on catastrophe theory, and things like that. Thom asked me to give a lecture at this seminar on the Hopf bifurcation. At that time, this bifurcation was generally unknown, but I had just started and Thom was such a famous person. So I said, of course, I will do that, but can you tell me what it is? He said, well it is not so well known though Hopf published this paper in '42, and of course, he lived in Germany and at that time. The mail didn't work so well at that time for some reason which I don't remember. So he gave me a paper and said, "Well, Hopf did it in an analytic way, try to do it more geometric." So I tried to do that, I gave a lecture, and then Ruelle came up to me and said, "Do you know why Hopf was interested in this?" And I said, "No." And now I refer back to the previous speaker and say that Hopf's idea was the bifurcation from one frequency to second frequency, and so on. Well, I have also followed in that period a course on hyperbolic dynamics by Mike Shub in his version of French [laughter], and so a new ideology of talking about being generic and being non-generic, especially being non-generic, started to be considered. So I said, "Well, but on a two torus, having parallel flow is non-generic and on the three torus, you probably can have quite a bit more." So then it was Ruelle who said, "Well, this is important. Even if this idea about turbulence is wrong, it is important to publish it." So that was the paper where we first introduced the word "strange attractor." I don't know and he doesn't know either who was the first one to invent that word and, hence, we should share the responsibility. The continuation of that story is what you have heard by the previous speaker. It finally got noticed by physicists, and finally, even noticed that this had been already discovered before by Lorenz. Well, that was my way of getting entangled by chaos. It is a pity that Ruelle, who has been mentioned already twice, is not here. [Applause]

James Yorke – Other volunteers. Peter Grassberger.

Peter Grassberger (Department of Physics and Astronomy, University of Calgary, Canada) – Before I entered university, I didn't read books by Norbert Wiener, but I read books by Schrödinger, Heisenberg, and others as such. My idea, when I entered university, was that you have to solve the great riddles of the

universe, and everything else is just between stamp collecting and measuring the viscosity of, well, whatever. So I never contemplated doing anything else, but I was interested in most things. So I tried very hard to work in particle physics, but it was tough with these very brilliant guys and with these very few ideas. Whenever there was a new idea, even the third or maybe the fifth idea on the market, then immediately people jumped on it and I had to run after them. It was frustrating. But then I had an idea about something I could do and which was different. I do not want to go into the details, but anyhow, I learned that sometimes there are some leads. You can just follow a path and you just have to let yourself go. Then you end up in a different field. Once I sneaked out of particle physics that way, I felt suddenly, "Oh! The whole world has opened!". Before I had done this, just a few months before, I was in Nice. There was this guy who was supposed to have done a very tough work on renormalization, really field theory, you know, many pages of calculations. But he was good-looking, he liked the girls . . . In Nice, you have the beach ... Suddenly some seminars by him were announced. And these seminars were about the parabola. You take the parabola, iterate it and then, you get renormalization and ..., you can guess. At that moment, I didn't realize that this was something for me. But it was fascinating, weird, completely different from the serious stuff a German professor was supposed to work on. It was a very unusual thing. Then later, after I had this experience that I could break out of my cage, I said, "Oh, God!". I freed up, came back and read Mandelbrot's book, which was also weird. I had this idea that you can measure fractal dimensions, and after that, there were some other ideas, and in that way I was dragged in. First we did something about the Feigenbaum attractor and this idea came: "Why don't we measure the dimension of strange attractors?" And then we came up with further ideas. Okay, that's it. I should say one more thing: in the life of every man a moment comes where chaos takes over. I have the luck that I can do it professionally. [Applause]

Edward Ott (Institute for Research in Electronics and Applied Physics, University of Maryland, USA) – Chaos is also called deterministic randomness and is associated with the difficulty of forecasting the future. My entry into chaos was chaotic in that sense. So, how did I get into chaos? Well, I was working in plasma physics, and in the United States you can't survive as an academic researcher unless you have contracts. And when you have a contract, you have a contract monitor. I had a contract and a contract monitor. One day I was going to a scientific meeting and I got on the plane, and by random chance, I was assigned a seat next to my contract monitor. His name was Oscar Manley, and he was reading a paper by Mikhail Rabinovich. I asked him what it was on, and he told me that it was on a strange attractor. As Takens in his remarks said, that is sort of a bewitching term. If you've never heard it before, it sounds really exotic. And indeed, as Oscar described it, it did sound like something I should get to know. So I eventually started to read that paper, and that led to my first paper on dissipative chaos. (I had been working on Hamiltonian chaos before that.) Another examples of randomness in my career in getting started in chaos came when I moved from Cornell, where I was for 11 years, to the University of Maryland. At the University of Maryland, we had a Physics

Department which had undergone rapid expansion. It had expanded so much that there was no room for me as a new faculty member in the Physics Building. So they put me in some other building off in a corner of the campus. It wasn't a building that just had physicists in it; it also had people like Jim Yorke, whose office was very close to mine. And we started talking. We would talk at the end of almost every day, and we decided that we should collaborate in research. So, my associating with Jim was another aspect of randomness. We decided we would try to get some contract money to do the research that we were thinking of. We submitted a proposal to Oscar Manley at DOE, and it was funded. One of the things the proposal provided for was a postdoc to work with us. And so we were saying, okay, let's hire our postdoc. But at the same time, there was a plasma physicist at the University of Maryland who had hired Celso Grebogi to work with him. Celso was about to come, and this fellow came to me and said, "Hey, I hired this guy but now I realize I have no money to pay him." [Laughter] "So could you help me out?" And so we hired Celso and that's how our chaos group got started. Thus my start in dissipative chaos research is an example of chaos and randomness where the associated random events have far reaching effects that are very unpredictable. [Applause]

James Yorke – About 80 papers completed, among the three of us. [Applause]

Lou Pecora (Naval Research Laboratory, Washington, USA) – Jim asked me earlier to speak to you, and I told him, "Well, I got interested because I took a course," and he said, "That's really boring. I don't want to have you stand up there and talk about that." Then after talking to some people at the table, I thought that I could tell you about how I got involved in synchronising chaotic systems, because I think that is an interesting story. So, where I work is the Naval Research Lab. To do something there, it is not good enough if it is interesting. They want to see you do something that can be applied, "Why are you doing that and what is it good for?" So Tom Carroll and I, Tom's my colleague, we said, "Well this really interesting stuff – nonlinear dynamics, chaos – what could we do with this?" One of the first things we thought about, is "Wouldn't it be nice if you could synchronise two chaotic systems?". You could maybe communicate with them or do something like that. We thought that if we could do that, it would be great, but we didn't have a clue of how to do it. And so, we spent a good part of the month of December really knocking down each other's ideas on how to do this. We went to, at that point, a dynamics-based conference in Houston in January, and came home. I remember being exhausted. I came home and my daughter, who was about 6 months old, woke up in the middle of the night. I couldn't sleep. I told my wife, "I will get up, I'll feed her." And I am sitting there feeding her and I'm thinking, "You know, you could drive a Duffing with a sine wave and put the sine wave in the Duffing, and you could drive a Lorenz with a Lorenz." I just remember being so tired and thinking, "I hope I remember that tomorrow morning." I put my daughter down and went to sleep. I should have written it down. But I did remember it. At that point, the only thing I could do was to say "This is an interesting idea. Drive one system with another." I did a couple of logistic maps and they worked. We proposed it at the Naval Research Lab and

they liked it. They had no idea what we were talking about, but it looked sexy and sometimes sexy wins. Then I realized that I had got to solve some differential equations at this point because we had really got to show this is for real chaotic systems like Rössler and Lorenz. So, I thought, "Next thing, I go into the lab, now that we have the money to do this, and I am going to do that." Except that I have three kids and my oldest son came down with chickenpox. And if you don't know what chickenpox is – it's probably something else in another language. It's a childhood disease that's usually not fatal but you gotta stay home. My wife had to work, so I stayed home. This was before we had computers in our house and I stayed there thinking, "I've got to get in there and do this." Human psychology is a funny thing. As soon as I got the idea, the first thing I thought of was, "Oh my God, someone is going to do it before me." Of course, I hadn't thought of that until now, but you suddenly get panicky when this happens because you are sure somebody is going to beat you to the punch. So I stayed home for 2 weeks with my son. He got better, I started to go in, and my second son got sick. Finally, I ended up having my daughter also getting sick. All three kids got sick. I spent 2 months at home and I could not get started. I was worried somebody was going to scoop us. That was my own problem. Nonetheless, our paper came out and I felt good about it because we had beaten everybody. Only later did I realize that we weren't the first people to synchronise chaotic systems. This is a kind of interesting. Somebody said today, "You think you did it first but if you look at the literature, which you probably should have done a little bit more, you find somebody else that had done something before you." So I think our paper was good because we suggested applications and I think that's why it's done well. But before us–our paper was published in 1990 – there was a paper in 1989 by Rul'kov et al. on synchronising chaotic systems, except they didn't suggest a use for it. But that's a Russian that had beat us. But before that, there was a very nice paper by Valentin Afraimovich, Misha Rabinovich and Verichev in 1986 synchronising stochastic systems, basically synchronising two Duffing systems. And in true Russian form, they also built a circuit like we did. The Russian theorists also can build circuits, which is really good. So we found this out later and I found out also that Arkady Pikovsky, who is here, in 1984, published *Interaction of Two Chaotic Attractors* and strange attractors, which was basically synchronising chaotic systems. So if you're a Russian now, you're feeling really good because the Russians have scooped everybody. But I actually found in 1983 there were three papers by Japanese researchers, Fujisaka and Yamada, who had synchronised chaotic systems. That's the earliest I know of. So I was already scooped and I didn't know it. This is kind of a lesson maybe to young people. Even if you are scooped, sometimes you can make something out of it. So that's not an introduction to chaos, but it is funny how things can just work out pretty well, and in the end that paper worked out pretty well for us. [Applause]

Itamar Procaccia (Department of Chemical Physics, The Weizmann Institute of Science, Rehovot, Israel) – So if you just squeeze my hand, I will take my revenge out and tell you what happened to the first paper that we sent for publication. So this was when Peter Grassberger visited me for a few months

for sabbatical and we wrote this little letter with a program of about five lines that described how to find the correlation dimension of a strange attractor. And we sent it for publication and it was rejected by two referees, who said this makes no sense. We got it back and the reasons were very clear. Apparently, we were a little bit late on the stage and people already knew everything. So we got two papers that had been published a couple months before. One was, I think, a paper by the Santa Cruz crowd, Farmer, etc., those who had determined that the dimension cannot be computed in finite time. These were the Santa Cruz people. It was new. But there was another paper, by Jim Yorke and I can't remember who else, saying that there are only two types of dimensions – the box dimension and the information dimension, and nothing else can exist. And then, we came up with a dimension we called the correlation dimension, which was obviously different. It was obviously a lower bound to both. So the paper was rejected. I remember that it was really quite a struggle. At some point, one referee said that since it had been shown that it cannot be like that, then we must be cheating if we get the result that we claim. Then the idea was: "Let's send them the program."

James Yorke – You should understand that the referee might have been interpreting what he said perhaps incorrectly and then imposing it upon him.

Itamar Procaccia – I will consider this as an explanation. And you want to insinuate that you were not the referee, I understand. [Laughter] Anyway, the end was that we had to send the program in. We sent the program in, it was five lines long or something like this, and asked the referee "Please, run the program yourself and see that you get this number". And that is how it went. I think this was my first example of writing a paper that was rejected immediately and then it had about 2,000 citations. But it was not the last time. I think by now I probably have enough statistics to claim that all the best papers cannot be published without a lengthy struggle. And you know Jim, thank you for being the referee or not being the referee. [Laughter]

James Yorke – In the face of history of chaos, a large amount of the basic papers were rejected, at least the first time around, partly because they were sometimes only half-baked, but also because they didn't follow the usual trend. People wanted only the best papers, which meant slight changes in most of the literature.

Michael Berry (H H Wills Physics Laboratory, Bristol, UK) – The organisers of this delightful celebration for Celso have persuaded me to reminisce about my involvement with the beginnings of what has come to be called "quantum chaos" (though since that phenomenon is hard to find I much prefer the term "quantum chaology", which describes what we really do).

In 1972, Kate Mount and I wrote a review article on semiclassical approximations–what I now call prehistoric semiclassical mechanics: mostly WKB theory. Towards the end of the writing, I encountered papers by Martin Gutzwiller and was tremendously impressed by them and praised them in the review because he

explicitly considered non-integrable systems. I don't think Gutzwiller had thought about chaos at that time and it wasn't mentioned in our review.

The publishers sent our article to be reviewed by Ian Percival. When I met him, he told me that he had just spent a year in the USA where he had encountered the work of "a wonderful Russian mathematician called Vladimir Arnol'd who is making very spectacular discoveries about a phenomenon that underlies those papers by Gutzwiller that you feature in your article. It's called chaos, and you should learn about it." Percival had already come to the view that associated with chaos there should be a different kind of quantum state. He called it the irregular spectrum, and published a paper about it in 1973–the first paper explicitly about quantum chaology, for which he deserves credit. It had a big influence on me, though not for several years.

Then I encountered the writings of Joseph Ford: a great man, a passionate man, an infuriating man, a lovely man. What a pity Joe is no longer with us. He called himself the evangelist of chaos and that is a very good description. From his papers I finally and fully realized the importance of what had been achieved by Arnold and also Yakov Sinai and Boris Chirikov in classical chaos, and determined to understand the impact this would have on quantum mechanics, that Percival had alerted me to. The best way to learn a new subject is to give lectures about it, so I gave a graduate lecture course in Bristol, based on the papers of those pioneers.

Attending my lectures was Michael Tabor, an intellectual refugee from the wastes of theoretical chemistry (as he and I put it in those days). He was very excited and wanted to work with me. But I told him: "I don't have much time, because we've just bought a new house, and I've promised to decorate it." A few weeks after moving in, there was a knock on the door. It was Tabor: "I've come to put my paint brush where my mouth is." So we spent the summer up a ladder, painting and thinking about chaos and quantum physics and we published three papers.

Meanwhile I was bold enough to send the notes for my lectures to Joe Ford. I was overwhelmed when he called me from Georgia; in those days, an international phone call was fairly unusual. In his wonderful southern accent, he arranged for the notes to be published in the proceedings of a meeting I had not attended, and invited me to the first international meeting on chaos, that he was organising with Giulio Casati in Como. This was 1977; I went there with Tabor.

At the meeting, I had the idea that the quantum stationary states of classically chaotic systems would have the structure of Gaussian random functions with a local coarse-grained intensity given by the classical microcanonical ensemble. André Voros was also at the meeting, and had a similar idea, but we published separately. Later, we learned that Alexander Shnirelman had published the microcanonical part of the idea a few years earlier.

A year before the Como meeting, Tabor and I had understood that random-matrix theory should give a good description of the statistics of energy levels in the irregular spectrum. This was as a result of the suggestion from my colleague Balazs Gyorffy that this subject–until regarded as a branch of nuclear physics, might be relevant to chaos. George Zaslavsky had had a similar idea in 1974, but unfortunately he did not identify the central feature; universality. Tabor and I published a brief account

of the envisaged random-matrix connection in one of our papers. This was taken up by Steven McDonald and Alan Kaufman, who we then met in Como; they made the first computations to check the idea (with the stadium quantum billiard). About 7 years later, the definitive and very influential statement of the connection was made, and supported by more accurate numerics, by Oriol Bohigas, Marie-Joya Giannoni, and Charles Schmidt, who knew much more than me about random matrices.

In a separate development, the head of the theory group in Bristol, John Ziman, a condensed-matter physicist with whom I never worked but who was something of a mentor to me, had shown me the papers on period-doubling by Robert May and by Jim Yorke. In Como I met Mitchell Feigenbaum, who made such a spectacular contribution by explaining the universality of these phenomena in terms of renormalization. My encounter with Giulio Casati soon led to discussions about the Aharonov-Bohm effect, and thus, indirectly, through connections I will not elaborate here, to the geometric phase in 1983. That Como meeting was unforgettable!

Ziman also showed me the book by René Thom on Catastrophe theory, thereby inspiring another major chapter in my scientific life (universal diffraction decoration of waves near caustics), but that too is another story.

Several years later, Alfredo Ozorio de Almeida, who had been my undergraduate student in the 1960s and my graduate student in the early 1970s, and my colleague John Hannay, made a major discovery about universality in the distribution of the long periodic orbits that appear in Gutzwiller's representation of the irregular spectrum. This became the basis of my 1985 understanding of the origin of random-matrix universality, and, equally important, how random-matrix theory fails for correlations between distant energy levels–an understanding that has been so ably developed and deepened by Jonathan Keating and Eugene Bogomolny, and most recently by Martin Sieber and Fritz Haake and others.

It was in the early 1980s (I do not remember the year) that I visited Maryland and first met the very creative classical chaos group of Jim Yorke, Edward Ott, and of course Celso Grebogi, who soon became a friend. I congratulate him today on reaching sixty: no younger person's age has more distinct prime factors. [Applause]

Giulio Casati (Center for Nonlinear and Complex Systems, Universitá dell' Insubria, Como, Italy) – My story is very simple and very short. My interest in chaos started when I just graduated, back to 1971. As a young Italian boy, I decided to look at the first work of Fermi, and the last work of his life. I don't know if you are aware of it, but they both refer to what is now called nonlinear dynamics. The first work is difficult to understand and is about quasi-ergodic theory. The last work is much more simple, even though profound, and it is the famous Fermi-Pasta-Ulam model. There were also two other persons most influential to me: one was Joe Ford, already mentioned by Michael Berry, and another one in Siberia, Boris Chirikov. This is how I entered the field. I just want to say that Chirikov is the person to be remembered since he has been one of the main pioneers in nonlinear dynamics and chaos. He has been also one of the last representative of the old 'Russian intellighenzia'. Thank you Mike for remembering him. [Applause]

Valentin Afraimovich (Universidad Autónoma de San Luis Potosí, Mexico) – I am Valentin Afraimovich from Russia, and in Russia we were surrounded by chaos. A former student from Andronov, N. N. Bautin, told us a story that he (Bautin) was trying to get a grant from the Ministry to work on a clock pendulum. He waited for long time for some burocrats at the Ministry. Over the wall there were some plaster mouldings that oscillated. He started to calculate how many oscillations per minute he observed. It was non-periodic behaviour. This was 1958. After that, he came home and created models of the dynamics of this device, which is called Gipp's pendulum. From the beginning we knew that there were some man-made devices which behaved non-periodically. That helped me very much when I started to do mathematical analysis of chaotic oscillations. [Applause]

Robert MacKay (Mathematics Institute, University of Warwick, UK) – Hi. I think I first encountered the theory of chaos after my second year as an undergraduate. My father pointed me to the now famous paper by Robert May in *Nature* in '76. I thought it was cute, interesting. The next encounter with chaos was in my final year (Part III) when I had to do an essay, and one of the topics was "Period three implies chaos". But I thought this didn't look serious enough material, so I did something else (waves in inhomogeneous atmospheres). Sorry Jim. Then the third encounter with chaos was as a beginning PhD student in Princeton Plasma Physics Lab in '78. We each did an experimental project in the first year. The person I was assigned to proposed we look at drift waves in a device called QED (quiet energetic dense). You destabilise them by turning up the magnetic field because that reduces the Larmor radius, and it produced a beautiful periodic spectrum. But I wanted to see what would happen if we turned up the field further, and what I found was a 2-frequency quasiperiodic spectrum. I turned even further and found frequency-locking and turned up even further, and found a broadband spectrum. We didn't know about the work of Swinney. Anyway, here was the two-frequency route to chaos under my eyes, but I didn't interpret it that way at the time, and my advisor didn't know what to make of it and switched me to something else (verifying the dispersion relation for whistler waves). The fourth encounter with chaos was in various graduate courses. Princeton was a fantastic place to be a graduate student. I had courses from John Krommes and Arthur Wightman, and people like that, containing aspects of Hamiltonian chaos theory (and applications like chaotic advection well before Aref's famous paper). And then a bunch of us graduate students discovered that we were all interested in nonlinear dynamics. So we formed a reading group and read up on some papers. One week, we drove my car to the New York Academy of Sciences conference on Nonlinear Dynamics in December '79 and wow, I hardly understood a thing but I was fascinated and felt this was the area I would like to work in. Then, after that, we decided to go to John Mather's course, where I learnt many topics in dynamical systems theory over two and half years, culminating in what is now called Aubry-Mather theory. But before we got to that topic, I heard a seminar by John Greene in Spring '80 on his overlap criterion for the breakup of tori, and I thought this was fascinating. So I asked if he had anything similar for me to work on, and he said yes; he was playing around with period doubling in area-preserving

maps, and invited me to join in. That was my beginning in research in chaos. But I would like to add two other encounters. Firstly, I took a practical course in electronics in Spring '79 and sometime in the next 2 years made an electronic circuit to demonstrate period-doubling sequences and chaos, like Chua's circuit, except that I made mine before Chua had invented his. I made another one to demonstrate the 2-frequency route to chaos, and demonstrated both at a PPL open day in November '81 under the banner "Bifurcations Live!". Secondly, I was fortunate to be accepted for the '81 Les Houches summer school on Deterministic Chaos, where I learnt a huge amount and encountered my future wife. Thanks. [Applause]

James Yorke – OK, changing gears. I do want to make a couple of remarks about Celso. Celso, not about the scientist, but about his personality. Everybody knows about him as a great experimenter of foods. He'll try anything. He'll come to you and say, "Here try this, you gotta try this!" And you try it, and you say, "Yah, this is terrible!" He says, "Yes, isn't it!" [Laughter] I wrote down in my computer an event that occurred 10 years ago. After Celso had been in the United States for over 15 years; 20 days from now it will be exactly 10 year when Celso experienced this fine American food, Diet Coke. [Laughter] Now I hate to think of the idea that he will be in Scotland without experiencing the great products of Great Britain. So I have brought him one ... Celso, where are you? Come on up. What we have here is a great product from England called a Yorkie bar. [Laughter] And I didn't want him to miss out on Yorkie bars. So I have brought three; one for you, one for your son, and one for your wife, because none of you should miss out on this wonderful product. [Applause] At this time, I hand the proceedings over to Celso.

Itamar Procaccia – Please, I have to say something, I have to say something.

James Yorke – Please.

Itamar Procaccia – This of course reminds me of this wonderful meeting in Italy, in Como, when we went together. And as you know, Jim is extremely American. So he comes to me and he asks me, "Is the water safe in this country?" [Laughter] "No, absolutely not! Don't touch it!" Then I had the pleasure of seeing you walk in with a Diet Coke for breakfast, lunch, and dinner. [Laughter]

James Yorke – I wrote my first paper on chaos when Andy Lasota, a Polish mathematician, visited Maryland. We worked on many projects. One project was on one-dimensional maps which were expanding. And he had the whole theory worked out on how to do it, except he couldn't quite make it work. So I was able to do the algebra to make it work and we got our first paper in 1973 on expanding maps and chaos. So that was my first chaos paper.

Celso Grebogi (Institute for Complex Systems and Mathematical Biology, University of Aberdeen, UK) – I wrote something that I should say tonight, but now it does not sound it right in view of how this evening is unfolding. Jim, what is

so typical of him, had this wonderful idea for us to recount how we did start working on chaos. It is proper for me then to relate too my first encounter with chaos. I went to Berkeley as a post-doc in September 1978 to work with Allan Kaufman. As a side comment, I will be back in Berkeley in 2 weeks for the Kaufmanfest to commemorate Allan's 80th. I should also mention that I am deeply grateful for Allan and his wife Louise for taking the time and making the effort to come to Aberdeen to be with us. Back to chaos, when I got in Berkeley in 1978, Allan has just received from Boris Chirikov an advanced copy of his path-breaking paper which was eventually published in Physics Reports in 1979. In order to go over the paper, Allan set up round table discussion sessions that met every Thursday, for the whole afternoon. The people present were Allan's group, his former associates–some of them were then working at Lawrence Livermore Lab., and some people from the Physics and Math. Departments. As a curiosity, often a bright and young looking fellow would come to the meetings and would answer all the difficult mathematical aspects of the paper; soon we found out that he was Alan Weinstein. Oscar Lanford was another person that often helped us with the mathematics of the theory of dynamical systems, especially the ergodic theory aspects. Those sessions were later enlarged with the inclusion of Jerrold Marsden, and led to the discussion and development of major projects in dynamics, especially the conservative one. Those were wonderful years in which I learned some of the mathematics of chaotic dynamics; I spent 3 years in Berkeley before moving to Maryland. Ed Ott has already told you the story of when I got there and initiated the collaboration with him and Jim. It remains for me to express to each one of you how deeply gratified and honoured I am for your presence at this Conference.

Singular Perturbations of Complex Analytic Dynamical Systems

Robert L. Devaney

Abstract Our goal in this paper is to describe the dynamical behavior of singular perturbations of complex dynamical systems. Singular perturbations occur when a pole is introduced into the dynamics of a polynomial. In this paper, we consider the simplest possible case: start with the map $z \mapsto z^n$ where $n > 1$ and then add a pole at the origin. For simplicity, we consider the case of maps of the form $z^n + \lambda/z^n$. We then describe some of the many ways Sierpinski curve Julia sets arise in this family. We also give a classification of the dynamics on these sets and describe the intricate structure that occurs around the McMullen domain in the parameter plane for these maps. Finally, we discuss some of the major differences between the cases $n = 2$ and $n > 2$.

1 Introduction

Our goal in this paper is to describe the behavior of singularly perturbed complex analytic dynamical systems. By a singular perturbation we mean the following. Suppose we have a complex analytic map F_0 which, for simplicity, we will assume to be a polynomial of degree d. Consider the new map F_λ obtained by adding a pole at $a \in \mathbb{C}$ so that

$$F_\lambda(z) = F_0(z) + \frac{\lambda}{(z-a)^m}$$

where $\lambda \in \mathbb{C}$ is a parameter. The map F_λ is a singular perturbation of F_0 as soon as $\lambda \neq 0$ since the degree of the map changes from d to $d + m$. As a consequence, the dynamical behavior of F_λ is much richer than that of F_0, although some portions of the dynamics of F_0 persist depending upon the location of the pole a.

R.L. Devaney (✉)
Department of Mathematics, Boston University, Boston, MA 02215
e-mail: bob@bu.edu.

M. Thiel et al. (eds.), *Nonlinear Dynamics and Chaos: Advances and Perspectives,*
Understanding Complex Systems, DOI 10.1007/978-3-642-04629-2_2,

The reason for the interest in singular perturbations arises from Newton's method. Suppose we are applying Newton's method to find the roots of a family of polynomials P_λ which has a multiple root at, say, the parameter $\lambda = 0$. For example, consider the especially simple case of $P_\lambda(z) = z^2 + \lambda$. When $\lambda = 0$ this polynomial has a multiple root at 0 and the Newton iteration function is simply $N_0(z) = z/2$. However, when $\lambda \neq 0$, the Newton iteration function becomes

$$N_\lambda(z) = \frac{z^2 - \lambda}{2z} = \frac{z}{2} - \frac{\lambda}{2z}$$

and we see that, as in the family F_λ, the degree jumps as we move away from $\lambda = 0$. In addition, instead of just having a globally attracting fixed point at the origin, after the perturbation, the dynamical behavior of N_λ become much richer.

For simplicity, in this paper, we will consider the simplest possible case where $F_0(z) = z^n$ and $n \geq 2$. So the dynamics of F_0 are well understood. If $|z| < 1$, the orbit of z tends to the attracting fixed point at the origin under iteration of F_0. If $|z| > 1$, the orbit of z tends to ∞ which is also an attracting fixed point for F_0 in the Riemann sphere. On the circle $|z| = 1$, the map is well known to be chaotic; this set is the *Julia set* of F_0, which will be defined below.

For most of this paper, the singularly perturbed family will be of the form

$$F_\lambda(z) = z^n + \frac{\lambda}{z^n}$$

where again $n \geq 2$. So the attracting fixed point at the origin is replaced by a pole, just as in the Newton's method example above.

We remark that we could equally well consider the case where the degree of perturbation is $d \neq n$, but the case $d = n$ has some special symmetries that make this case easier to understand.

The principal goal of this paper is to give a survey of some of the many recent results concerning the dynamical and parameter planes for the family F_λ. A subtheme in this paper is to show that the case where $n = 2$ is quite different from the case where $n > 2$. Indeed, this lower degree family of maps turns out to have much more complicated dynamical behavior than the case where $n > 2$. Another subtheme is to describe some of the interesting topological spaces that arise as the Julia sets of the maps F_λ.

2 Preliminaries

Consider the family of maps

$$F_\lambda(z) = z^n + \frac{\lambda}{z^n}$$

where $n > 2$. The *Julia set* of F_λ, denoted by $J(F_\lambda)$, is defined to be the set of points at which the family of iterates of F_λ fails to be a normal family in the sense

of Montel. There are many other equivalent definitions of the Julia set. For example, the Julia set is the closure of the set of repelling periodic points of F_λ, and, in our special case, it is also the boundary of the set of points whose orbits escape to ∞. As a consequence, the Julia set is the set of points on which F_λ behaves chaotically, since arbitrarily close to any point in $J(F_\lambda)$ there is both a repelling periodic point and a point whose orbit escapes to ∞. The complement of the Julia set is called the *Fatou set*.

There are a number of critical points for F_λ. One critical point occurs at ∞, which is always a fixed point. A second critical point occurs at the origin, which is mapped directly to ∞. A straightforward computation shows that there are $2n$ other critical points for F_λ given by $c_\lambda = \lambda^{1/2n}$. These are the *free* critical points for F_λ. Despite the fact that there are $2n$ free critical points, there are only two critical values given by $v_\lambda = \pm 2\sqrt{\lambda}$. However, there really is only one free critical orbit since, when n is even, both of the critical values are mapped to the same point. When n is odd, the orbits of the two critical values behave symmetrically under the map $z \mapsto -z$. Thus this family of maps, like the quadratic polynomial family, is a natural one-parameter family of maps. The parameter plane (the λ-plane) is then a record of the behavior of the free critical orbit, just as in the case of the Mandelbrot set. There are also $2n$ prepoles at the points $p_\lambda = (-\lambda)^{1/2n}$, i.e., $F_\lambda(p_\lambda) = 0$.

Here is one reason why the case $n = 2$ is so different from the case $n > 2$. Consider the second iterate of the critical points. We compute

$$F_\lambda^2(c_\lambda) = 2^n \lambda^{n/2} + \frac{1}{2^n \lambda^{(n/2)-1}}.$$

When $n > 2$, it follows that $F_\lambda^2(c_\lambda) \to \infty$ as $\lambda \to 0$. But when $n = 2$, the second iterate of c_λ reduces to $4\lambda + 1/4$, so $F_\lambda^2(c_\lambda) \to 1/4$ as $\lambda \to 0$. The reason this is significant will become clear in the next section.

Let C_λ be the circle given by $|z| = |\lambda|^{1/2n}$. Note that C_λ contains all of the critical points and the prepoles. A straightforward computation shows that F_λ maps C_λ $2n$ to 1 onto the straight line connecting the two critical values. We call C_λ the *critical circle* and its image the *critical line*. This is not the case for the family $z^n + \lambda/z^d$ with $n \neq d$, so this is one of the reasons why we deal only with the case $n = d$. One may also check that any other circle centered at the origin is mapped n to 1 onto an ellipse whose foci are $\pm v_\lambda$. Also, the straight ray connecting 0 to ∞ and passing through c_λ (a *critical ray*) is mapped 2 to 1 onto the straight ray connecting either v_λ or $-v_\lambda$ to ∞ and extending the critical line. Similarly, each straight ray connecting 0 to ∞ and passing through p_λ (a *prepole ray*) is mapped 1 to 1 onto the straight line passing through the origin and perpendicular to the critical line.

The maps F_λ have additional symmetries. Let ω be a primitive $2n$th root of unity. Then we have that $F_\lambda(\omega z) = \omega^n F_\lambda(z)$. Hence the orbits of points of the form $\omega^j z$ all behave "symmetrically" under iteration of F_λ. For example, if $F_\lambda^i(z) \to \infty$, then $F_\lambda^i(\omega^k z)$ also tends to ∞ for each k. If $F_\lambda^i(z)$ tends to an attracting cycle, then so does $F_\lambda^i(\omega^k z)$. However, the cycles involved may be different depending on k and, indeed, they may even have different periods. Nonetheless, all points lying on this

set of attracting cycles are of the form $\omega^j z_0$ for some $z_0 \in \mathbb{C}$. Another symmetry is given by the involution $H_\lambda(z) = \sqrt{\lambda}/z$. Here we have $F_\lambda(H_\lambda(z)) = F_\lambda(z)$.

When $|z|$ is large, the term λ/z^n in the formula for F_λ is negligible, so $F_\lambda(z) \approx z^n$ near ∞. Consequently, the point at ∞ is a superattracting fixed point for F_λ and it is well known that F_λ is conjugate to $z \mapsto z^n$ in a neighborhood of ∞, so we have an immediate basin of attraction B_λ at ∞. Since F_λ has a pole of order n at 0, there is an open neighborhood of 0 that is mapped n to 1 onto a neighborhood of ∞ in B_λ. If the entire basin of ∞ is disjoint from this neighborhood around the origin, then there is a open set about 0 that is mapped n to 1 onto B_λ, and this entire set is disjoint from B_λ. This set is called the *trap door*, since any orbit that eventually enters B_λ must do so by passing through the trap door. We denote the trap door by T_λ.

Using the symmetry $F_\lambda(\omega z) = \omega^n F_\lambda(z)$, it is straightforward to check that B_λ, T_λ, and $J(F_\lambda)$ are all symmetric under $z \mapsto \omega z$. We say that these sets possess $2n$-fold symmetry. In particular, since the critical points are arranged symmetrically about the origin, it follows that if one of the critical points lies in B_λ (resp., T_λ), then all of the critical points lie in B_λ (resp., T_λ). Also, the map H_λ interchanges B_λ and T_λ and preserves both the Julia and the Fatou set.

For other components of the Fatou set, the symmetry situation is somewhat different: either a component contains $\omega^j z_0$ for a given z_0 in the Fatou set and all $j \in \mathbb{Z}$, or else such a component contains none of the $\omega^j z_0$ with $j \neq 0 \bmod 2n$. See [1] for a proof of this fact.

3 The Escape Trichotomy

For the well-studied family of quadratic maps $Q_c(z) = z^2 + c$ with c a complex parameter there is the well known Fundamental Dichotomy:

1. If the orbit of the only free critical point 0 tends to ∞, then the Julia set of Q_c is a Cantor set;
2. If the orbit of 0 does not tend to ∞, then the Julia set is a connected set.

In this section we discuss a similar result for F_λ that we call the Escape Trichotomy. Unlike the family of quadratic maps Q_c, there exist three different "ways" that the critical orbit for F_λ can tend to infinity. If the critical orbit tends to infinity, then all of the critical values must lie in either B_λ, T_λ, or some of preimages of T_λ. These three different possibilities lead to the three distinct types of Julia sets for F_λ that comprise the Escape Trichotomy.

Theorem 1 (The Escape Trichotomy) *Suppose $n \geq 2$ and that the orbits of the free critical points of F_λ tend to ∞. Then:*

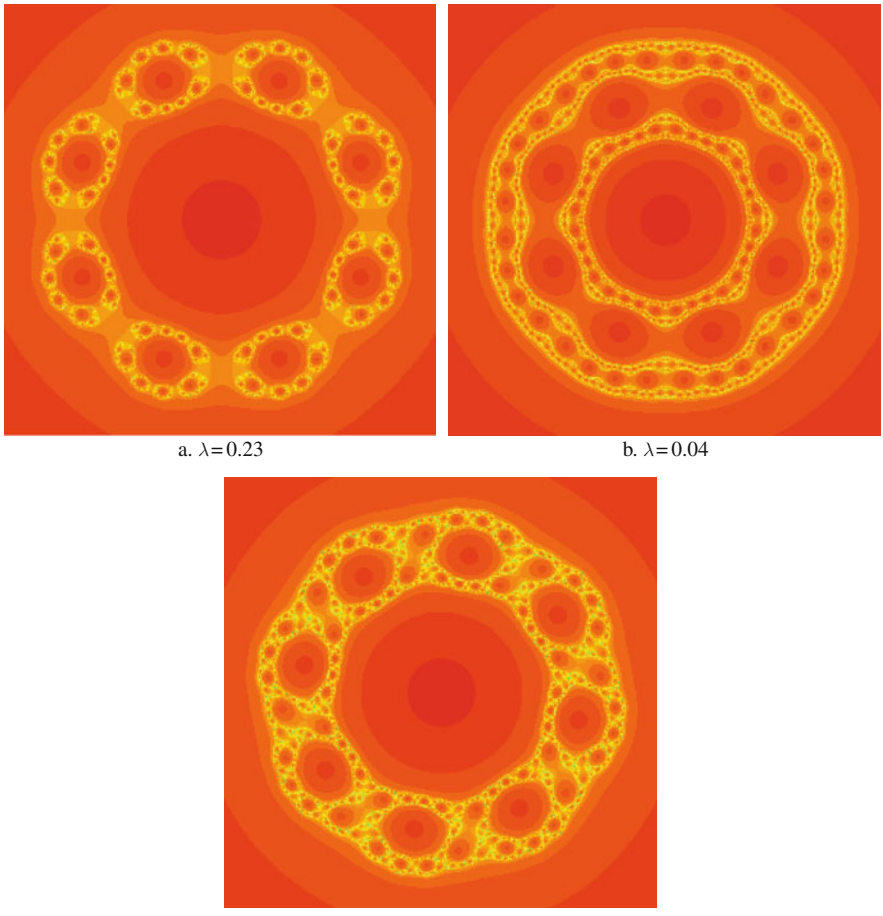
1. *If one of the critical values lies in B_λ, then $J(F_\lambda)$ is a Cantor set and $F_\lambda \mid J(F_\lambda)$ is a one-sided shift on $2n$ symbols. Otherwise, the preimage T_λ is disjoint from B_λ.*

2. *If one of the critical values lies in $T_\lambda \neq B_\lambda$, then $J(F_\lambda)$ is a Cantor set of simple closed curves (quasicircles), all concentric about the origin. This case does not occur when $n = 2$.*
3. *If one of the critical values lies in a preimage of B_λ that is different from T_λ, then $J(F_\lambda)$ is a Sierpinski curve.*

If the critical orbits never escape to ∞, then $J(F_\lambda)$ is a connected set.

Several Julia sets illustrating this trichotomy and drawn from the family where $n = 4$ are included in Fig. 1.

A *Sierpinski curve* is a very interesting topological space. By definition, a Sierpinski curve is a planar set that is homeomorphic to the well-known Sierpinski



a. $\lambda = 0.23$ b. $\lambda = 0.04$

c. $\lambda = -0.12$

Fig. 1 Some Julia sets for $z^4 + \lambda/z^4$: if $\lambda = 0.23$, $J(F_\lambda)$ is a Cantor set; if $\lambda = 0.0006$, $J(F_\lambda)$ is a Cantor set of *circles*; and if $\lambda = 0.125i$, $J(F_\lambda)$ is a Sierpinski curve

carpet fractal. But a Sierpinski curve has an alternative topological characterization: any planar set that is compact, connected, locally connected, nowhere dense, and has the property that any two complementary domains are bounded by disjoint simple closed curves is known to be homeomorphic to the Sierpinski carpet [2]. Moreover, such a set is a universal planar set in the sense that it contains a homeomorphic copy of any compact, connected, one-dimensional subset of the plane.

We remark that the second part of the Escape Trichotomy was first proved by McMullen [3]. He showed that this result only holds if $n > 2$ and λ is sufficiently close to 0. Indeed, the critical values can never lie in T_λ when $n = 2$ for $|\lambda|$ small since the image of v_λ tends to $1/4$ as $\lambda \to 0$. When $n > 2$, the image of v_λ tends ∞ as $\lambda \to 0$.

In Fig. 2, we show the λ plane in the case $n = 4$. The outer region in this image consists of λ-values for which $J(F_\lambda)$ is a Cantor set and is called the *Cantor locus*. The central disk is the *McMullen domain* in which $J(F_\lambda)$ is a Cantor set of simple closed curves. The region between these two sets is called the *connectedness locus* as the Julia sets are always connected when λ lies in this region. The other disks in the connectedness locus correspond to *Sierpinski holes* in which the corresponding Julia sets are Sierpinski curves.

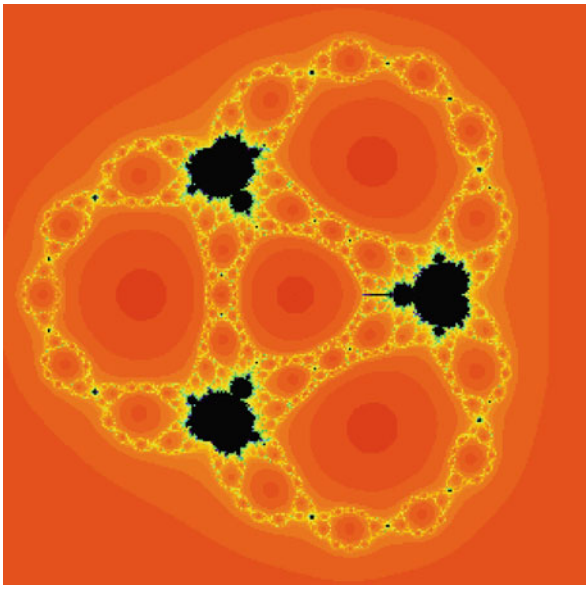


Fig. 2 The parameter plane when $n = 4$

4 Proof of the Escape Trichotomy

In this section we provide a rough sketch of the proof of the Escape Trichotomy Theorem.

First let v_λ be one of the critical values of F_λ and assume $v_\lambda \in B_\lambda$. So $\pm v_\lambda$ both lie in B_λ. Let $\pm\gamma$ be two symmetric and disjoint curves lying in B_λ and connecting

$\pm v_\lambda$ to ∞. Let μ_j be the $2n$ preimages of $\pm\gamma$ for $j = 1, \ldots, 2n$. One checks easily that each μ_j lies in B_λ, extends from 0 to ∞, and contains a single critical point. Hence B_λ and T_λ are not disjoint. Let U be a neighborhood of ∞ that contains $\pm\gamma$ and is $2n$-fold symmetric. We may choose U so that U meets each μ_j in a single arc. Let $V = H_\lambda(U)$ where we may assume that U and V are disjoint. Then both U and V lie in B_λ as do the μ_j. So $J(F_\lambda)$ lies in the complement of the union of U, V, and the μ_j, which is a collection of $2n$ simply connected sets, $I_1, \ldots, I_{2n}$. It follows easily that each of the I_j is mapped univalently onto a region that contains all of the I_k. Then standard techniques from complex dynamics show that the set of points whose orbits remain for all time in the union of the I_k is a Cantor set and the restriction of F_λ to this set is conjugate to the one-sided shift map on $2n$ symbols. This Cantor set is $J(F_\lambda)$.

For case 2, we have by assumption that B_λ and T_λ are disjoint open disks and all of the critical points lie in $F_\lambda^{-1}(T_\lambda)$. We claim that $F_\lambda^{-1}(T_\lambda)$ is an open annulus A. To see this note that $F_\lambda^{-1}(T_\lambda)$ cannot be a collection of $2n$ disjoint disks, for, in such a case, each of the disks would be mapped at least 2 to 1 onto T_λ. Therefore there would be at least $4n$ preimages of any point in T_λ, contradicting the fact that the degree of F_λ is $2n$. Hence at least two of the preimages of T_λ must intersect. But then, by symmetry, all of these preimages must intersect so the preimage of T_λ is a single open connected set. But then the Riemann – Hurwitz formula implies that A must be an annulus that contains all of the critical points.

Now A lies in the annular region that separates B_λ and T_λ and divides this region into two closed subannuli, A_{in} and A_{out}. Then F_λ takes each of these annuli as an n to 1 covering onto the entire complement of $B_\lambda \cup T_\lambda$. This implies that there are a pair of subannuli that are mapped onto A, one of which lies in each of A_{in} and the other in A_{out}. Continuing, one shows that $F_\lambda^{-j}(T_\lambda)$ consists of 2^{j-1} open, disjoint subannuli, each of which lies in the Fatou set. The complement of the union of all of these annuli (together with B_λ and T_λ) is then the Julia set which can be shown to be a Cantor set of simple closed curves.

Incidentally, the above argument indicates why this case cannot happen when $n = 2$. Since A_{in} and A_{out} are each mapped as 2 to 1 coverings of the annulus $\mathbb{C} - \{B_\lambda \cup T_\lambda\}$, it follows that the modulus of each of these subannuli is exactly half of the modulus of the bigger annulus. This implies that there is no room in $\mathbb{C} - \{B_\lambda \cup T_\lambda\}$ for any other sets, eliminating the possibility of the existence of A and further preimages of T_λ. So v_λ cannot lie in T_λ when $n = 2$.

We now describe the final case of the Escape Trichotomy where the critical values have orbits that eventually escape through the trap door, but the critical values themselves do not lie in T_λ. In this case the Julia set is a Sierpinski curve. To show this, we need to verify the five properties that characterize a Sierpinski curve. It turns out that four and a half of these preperties are trivial to show. First and second, since we are assuming that the critical orbit eventually enters the basin of ∞, we have that the Julia set is given by $\mathbb{C} - \cup F_\lambda^{-j}(B_\lambda)$. That is, $J(F_\lambda)$ is $\mathbb{C}$ with countably many disjoint, simply connected, open sets removed. Hence $J(F_\lambda)$ is compact and connected. Third, since $J(F_\lambda) \neq \mathbb{C}$, it is known that $J(F_\lambda)$ cannot contain any open sets, so $J(F_\lambda)$ is nowhere dense. Fourth, since the critical orbits all tend to ∞ and hence do not lie in or accumulate on $J(F_\lambda)$, it follows that F_λ is hyperbolic on $J(F_\lambda)$

and standard arguments show that $J(F_\lambda)$ is locally connected (see [4]). Hence $J(F_\lambda)$ fulfills the first four of the conditions to be a Sierpinski curve.

To finish proving that $J(F_\lambda)$ is a Sierpinski curve we need to show that the boundaries of B_λ as well as all of the preimages of B_λ are simple closed curves and that these boundary curves are pairwise disjoint. Now, since B_λ is a simply connected component of the Fatou set, it follows that the boundary of B_λ is locally connected. However, this boundary may have pinch points as in the case of quadratic Julia sets such as the basilica or Douady's rabbit. This is the only non-standard portion of the proof. However, assuming these boundaries are indeed simple closed curves, they cannot intersect because any such intersection point would necessarily be a critical point. This follows because, for some high power k, F_λ^k takes both of these curves to the boundary of B_λ, so the orbit of any intersection point must pass through a critical point since F_λ^k is locally 2 to 1 there. For the proof that these boundaries are simple closed curves, we refer to [1].

5 Classification of Escape Time Julia Sets

As can be seen in Fig. 2, there are a large number of Sierpinski holes in the parameter planes for these maps. We say that such a hole has *escape time* κ if, for each λ in the hole, the critical orbits land in B_λ at iteration κ. A parameter λ is called the center of the Sierpinski hole if the orbit of the critical points of F_λ all land on the point at ∞ rather than tend to ∞. The following result is proved in [5, 6].

Theorem 2 *There is a unique center of each Sierpinski hole. Moreover, there are exactly* $(n-1)(2n)^{\kappa-3}$ *Sierpinski holes with escape time* κ *in the parameter plane.*

The proof of this result uses quasiconformal surgery techniques to show that there is a unique center of each Sierpinski hole. Given this, the equation for the centers of the holes, namely $F_\lambda^{\kappa-1}(c_\lambda) = 0$, is easily seen to reduce to a polynomial equation of degree $(n-1)(2n)^{\kappa-3}$, and so the roots of this equation are all distinct.

As an example of the above count of Sierpinski holes, when $n = 3$ there are 2 Sierpinski holes in the parameter plane with escape time 3, 12 holes with escape time 4, 432 Sierpinski holes with escape time 6, and 120, 932, 352 holes with escape time 13. All of the parameters from this large collection of holes thus have Julia sets that are homeomorphic, so the natural question is: are the dynamics on these Julia sets the same?

The answer to this question is given in [7]:

Theorem 3 (Escape Time Conjugacy) *Let*

$$F_\lambda(z) = z^n + \frac{\lambda}{z^n} \text{ and } F_\mu(z) = z^n + \frac{\mu}{z^n}$$

where λ *and* μ *are parameters that lie in Sierpinski holes.*

1. *If* λ *and* μ *lie in the same Sierpinski hole, then* F_λ *and* F_μ *are topologically conjugate on their Julia sets;*

2. *If λ and μ lie in Sierpinski holes with different escape times, then F_λ and F_μ are not topologically conjugate on their Julia sets;*
3. *Suppose λ and μ are centers of different Sierpinski holes that have the same escape time. Let α be a primitive* $(n-1)$st *root of unity. Then F_λ and F_μ are topologically conjugate on their Julia sets if and only if, for some integer j,*

   - $\mu = \alpha^{2j}\lambda$, *or*
   - $\mu = \alpha^{2j}\overline{\lambda}$.

   Therefore, if λ and μ are parameters lying in different Sierpinski holes whose escape times are the same, then F_λ and F_μ are topologically conjugate on their Julia sets if and only if the centers of these Sierpinski holes have the above property.

The proof of the first part of this theorem follows by quasiconformal surgery techniques. The second part follows from the fact that any conjugacy between F_λ and F_μ must take ∂B_λ to ∂B_μ, ∂T_λ to ∂T_μ, and the kth preimages of ∂T_λ to the corresponding preimages of ∂T_μ. But the preimages of T_λ and T_μ containing the critical points are special: their boundaries are mapped 2 to 1 onto their images, and these are the only preimages of ∂T_λ and ∂T_μ that have this property. Hence, two such conjugate maps must have the same escape times. Finally, for part three, it suffices to consider the maps that are the centers of the corresponding holes. But these maps are "critically finite" in the sense that all of the critical orbits eventually land on the fixed point at ∞. By Thurston's Theorem [8], in the orientation preserving case, two such maps can be conjugated by a Möbius transformation. But such a conjugacy must then take ∞ to ∞ (since ∞ is the only superattracting fixed point) and 0 to 0 (since 0 is the only preimage of ∞). It follows that the conjugacy must be of the form $z \mapsto \alpha z$. Then, comparing coefficients in the conjugacy equation

$$\alpha F_\lambda(z) = F_\mu(\alpha z)$$

shows that $\alpha^{n-1} = 1$ and $\mu = \alpha^2\lambda$. In the case of an orientation reversing conjugacy, it is easy to check that F_λ is conjugate to $F_{\overline{\lambda}}$ via $z \mapsto \overline{z}$, so this gives all of the possible conjugate centers of Sierpinski holes.

This result allows us to give a precise count of the number of different conjugacy classes of escape time Sierpinski curves, because only those holes that are symmetric under rotation by successive squares of a primitive $(n-1)$st root of unity or by complex conjugation have the same dynamics.

Theorem 4 (Number of Conjugacy Classes) *The number of topological conjugacy classes of escape time Sierpinski curve Julia sets with escape time κ is given by*

1. $(2n)^{\kappa-3}$ *if n is odd;*
2. $(2n)^{\kappa-3}/2 + 2^{\kappa-4}$ *if n is even.*

For example, when $n = 3$, we have seen that there are exactly 432 Sierpinski holes in this family with escape time 6, but there are only 216 different conjugacy classes of such maps. Similarly, there are 120, 932, 352 Sierpinski holes with escape

time 13 but only 60, 466, 176 different conjugacy classes, so clearly there is a great variety of different dynamical behaviors on these escape time Sierpinski curve Julia sets.

The reason for the different number of conjugacy classes when n is even and odd comes from the fact that, when n is odd, there are no Sierpinski holes that meet the real axis (and so have no comple conjugate holes). Along the real axis there is only a pair of Mandelbrot sets and the McMullen domain. As a consequence, there are always exactly $n-1$ different Sierpinski holes with conjugate dynamics. See Fig. 3 for a picture of the parameter plane when $n = 3$. When n is even, the situation is very different; there is always a Cantor necklace along the negative real axis (more about this in Sect. 7). See Figs. 2 and 5 for pictures of the parameter planes when $n = 4$ and $n = 2$.

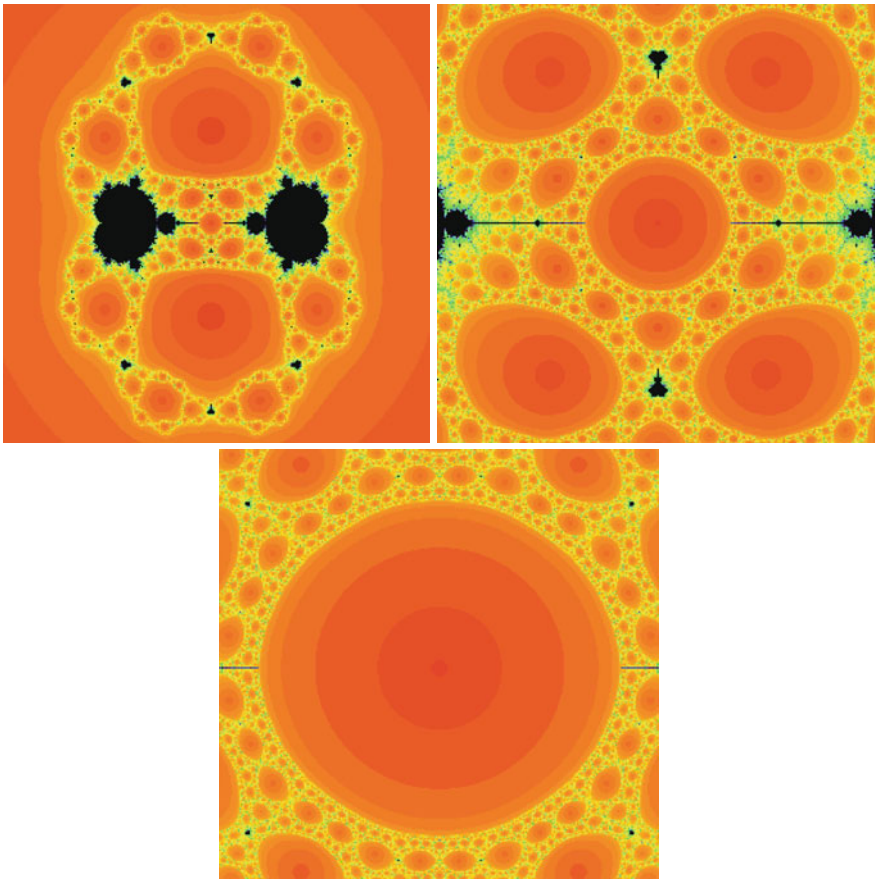


Fig. 3 The parameter plane for the family $z^3 + \lambda/z^3$ and several magnifications. The central disk is the McMullen domain $\mathcal{M}$. The simple closed *curves* in the Theorem accumulate on the boundary of $\mathcal{M}$

6 Structure Around the McMullen Domain

Recall that, when $n > 2$, if $|\lambda|$ is sufficiently small, v_λ lies in T_λ and $J(F_\lambda)$ is a Cantor set of simple closed curves. The entire region in the parameter plane for which this occurs is called the McMullen domain, $\mathcal{M}$. It is known [9] that $\mathcal{M}$ is an open, simply connected region that is bounded by a simple closed curve. In this section we describe some of the remarkable structure that surrounds $\mathcal{M}$ in the parameter plane. Since there is no McMullen domain when $n = 2$, the structure we describe below is absent in this case.

In Fig. 3 we display the parameter plane for $n = 3$ together with two successive magnifications around $\mathcal{M}$. Note that there appears to be a collection of closed curves surrounding $\partial\mathcal{M}$ that pass through more and more Sierpinski holes as these curves approach $\partial\mathcal{M}$. Closer inspection seems to indicate that these curves also pass through small copies of Mandelbrot sets as well. This is indeed true, as the following result was shown in [5, 10].

Theorem 5 (Rings around the McMullen domain) *For each $n \geq 3$, the McMullen domain is surrounded by infinitely many simple closed curves $\mathcal{S}^k$ for $k = 1,2,\ldots$ having the property that:*

1. *Each curve $\mathcal{S}^k$ surrounds $\mathcal{M}$ as well as $\mathcal{S}^{k+1}$, and the $\mathcal{S}^k$ accumulate on the boundary of the McMullen domain as $k \to \infty$.*
2. *The curve $\mathcal{S}^k$ meets the centers of τ_k^n Sierpinski holes, each with escape time $k + 2$, where*

$$\tau_k^n = (n-2)n^{k-1} + 1.$$

3. *The curve $\mathcal{S}^k$ also passes through τ_k^n centers of baby Mandelbrot sets, and these Mandelbrot sets and Sierpinski holes alternate as the parameter winds around $\mathcal{S}^k$.*

By a center of a baby Mandelbrot set, we mean the parameter drawn from the main cardioid of the Mandelbrot set for which the corresponding attracting cycle is actually superattracting, i.e., one of the critical points of F_λ is periodic. It turns out that all of these baby Mandelbrot sets are buried in the sense that they do not touch the outer boundary of the connectedness locus in the parameter plane. Then it is known that any parameter drawn from the main cardioid of this Mandelbrot set has a Julia set that is also a Sierpinski curve. These Sierpinski curves are dynamically very different from the escape time Sierpinski curves produced by the Escape Trichotomy, since each has an attracting cycle whose boundary curves are disjoint simple closed curves that are invariant under some iterate of F_λ. For an escape time Sierpinski curve, there is only one invariant boundary curve, namely ∂B_λ. For a proof that parameters from the main cardioids of buried baby Mandlebrot sets also yield Sierpinski curves, see [5, 11].

Here are some of the ideas involved in the proof of the rings around $\mathcal{M}$ theorem. When λ satisfies $|\lambda| < 2^{-2n/(n-1)}$, one checks easily that $|v_\lambda| < |c_\lambda|$. So the critical circle C_λ lies in the exterior of its image, the critical line. Then there is a preimage

of C_λ, ζ_λ^1, that lies outside of C_λ and is mapped n to 1 onto C_λ. Then there is an outer preimage of ζ_λ^1, ζ_λ^2, that is mapped n to 1 to ζ_λ^1, and so forth. We thus find an infinite collection of closed curves ζ_λ^k moving outward from the critical circle in the dynamical plane and having the property that each ζ_λ^k contains exactly $n^k \cdot 2n$ points that are mapped by F_λ^k to one of the critical points in C_λ and the same number of points that are similarly mapped to a prepole in C_λ. Now F_λ takes the interior of C_λ to the complement of the critical line as an n to 1 covering map. So one can consider the map $\phi(\lambda) = F_\lambda(v_\lambda)$. This is a map that takes the parameter plane to the dynamical plane. One can show that this map is univalent on each of the $n-1$ sectors in the parameter plane bounded by the straight rays through the "spines" of the $n-1$ large Mandelbrot sets symmetrically arranged around the origin. Moreover, ϕ takes each such sector onto $\mathbb{C}$ minus a pair of half-lines.

Now consider a particular kth preimage of one of the critical points lying in C_λ that lies in ζ_λ^k. Call this point u_λ. Then u_λ varies analytically with λ as λ ranges over each of the sectors. So we can consider the analytic map $\Phi(\lambda) = \phi^{-1}(u_\lambda)$. This map takes the sector in the parameter plane to itself. Then one can show using the Schwarz Lemma that Φ has a unique fixed point in this sector. This fixed point is a parameter λ^* for which $\phi(\lambda^*) = u_{\lambda^*}$, i.e., $F_{\lambda^*}(v_{\lambda^*})$ lands on the given kth preimage of a critical point. Then either this critical point or its negative is fixed by $F_{\lambda^*}^{k+2}$. This produces a center of a baby Mandelbrot set for each of the given critical points on ζ_λ^k (modulo an identification as λ winds around the origin). We similarly get centers of Sierpinski holes by letting u_λ be a preimage of a prepole instead of a critical point.

7 Cantor Necklaces

There is another structure called *Cantor necklaces* that occurs in both the dynamical and parameter planes for these maps. A Cantor necklace is a planar set that is homeomorphic to the following set. Consider the Cantor middle-thirds set lying in the unit interval. Replace each removed open interval with a circular open disk whose diameter is the same as the length of the removed interval and whose boundary touches the two endpoints of the removed interval. The union of the Cantor set with these countably many open disks is a Cantor necklace.

To see Cantor necklaces in these families of maps, we restrict for simplicity to the case $n = 2$ (though the same proof works more or less verbatim when $n > 2$). Let the four critical points of F_λ be denoted by $c_0, \ldots, c_3$ where the c_j vary analytically with λ and, when $\lambda \in \mathbb{R}^+$, $c_0 \in \mathbb{R}^+$ and the other critcal points are arranged around the origin in counterclockwise fashion. For each λ in the connectedness locus, we may pick $\nu > 1$ so that the circle of radius ν lies in B_λ. Let β_λ be the preimage of this circle that lies in B_λ and let τ_λ be the corresponding preimage in T_λ. Then consider the following two regions I_0 and I_1. The set I_0 is the smaller of the two regions bounded by the critical rays through c_0 and c_3 and the curves β_λ and τ_λ. Let $I_1 = -I_0$. Then F_λ maps both I_0 and I_1 univalently over $I_0 \cup I_1$. Standard techniques from complex dynamics then show that the set of points whose orbits remain for all time in $I_0 \cup I_1$ is a Cantor set, just as in part 1 of the Escape Trichotomy.

Now there is a fixed point in I_0 and a preimage of this fixed point in I_1 and one checks easily that both of these points lie in ∂B_λ. Similarly, there are two preimages of the prefixed point in ∂T_λ, and pre-preimages of these points in two preimages of ∂T_λ. Continuing in this manner, we may adjoin all of these appropriate preimages of T_λ to the "endpoints" in the Cantor set and the resulting set is a Cantor necklace.

To see at least a portion of the Cantor necklace in the parameter plane when $n = 2$, consider the region D in parameter plane that is the portion of the disk of radius 2 centered at the origin that lies in the left half plane. Since the second iterate of the critical point is given by the map $\phi(\lambda) = 4\lambda + 1/4$, it follows that $\phi(D)$ is a half disk that strictly contains D. Now, for each $\lambda \in D$, we can construct the Cantor set in the dynamical plane as above. Furthermore, if $\lambda \in D$, then the set I_1 is easily seen to be contained in the disk D for each $\lambda \in D$, provided the outer radius $\nu < 2$. (The Cantor set construction holds even if λ is not in the connectedness locus.)

Now fix any point z_λ in the Cantor set. Here z_λ is a point with a given itinerary in $I_0 \cup I_1$, so z_λ varies analytically with λ. Thus we have two maps that are defined on D, the map ϕ and z_λ. The map ϕ is invertible, so one may check easily that the map $\lambda \mapsto \phi^{-1}(z_\lambda)$ takes D inside itself. By the Schwarz Lemma, this map then has a unique fixed point in D. This point is the unique parameter for which the second iterate of the critical points all land on the given point in the Cantor set. Hence we get one such parameter corresponding to each point in the Cantor set, and these points then form the Cantor set portion of the Cantor necklace in the parameter plane. To get the other part of the necklace, we adjoin the associated Sierpinski holes just as we did in the dynamical plane case.

One may extend this necklace to include parameters for which the critical orbits land on the portion of the Cantor set in I_0. See [12] for details. Figure 4 displays the

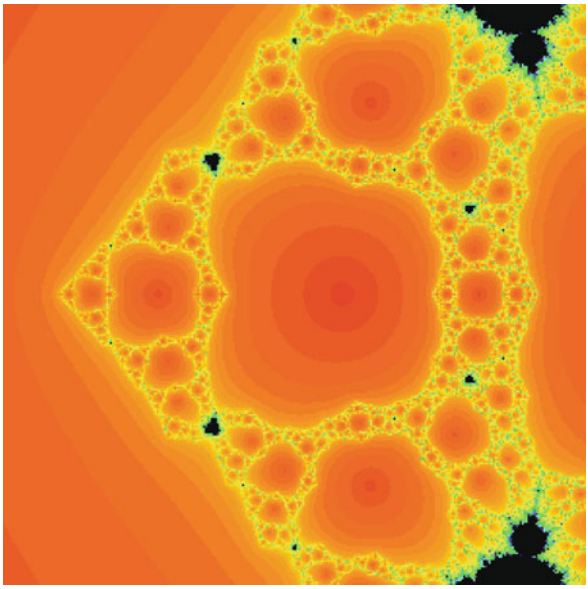


Fig. 4 A portion of the Cantor necklace in the parameter plane for the family $z^2 + \lambda/z^2$. Note the large Sierpinski hole along the negative real axis flanked by two smaller Sierpinski holes which are, in turn, each flanked by two smaller Sierpinski holes, etc

portion of the Cantor necklace corresponding to parameters for which the critical orbits land on points in I_1.

8 The Case *n*=2

As mentioned above, the situation in the case $n = 2$ is very different. We no longer have a McMullen domain. Rather, the following result is shown in [12]:

Theorem 6 *Suppose $n = 2$. Then, in every neighborhood of the origin in the parameter plane, there are infinitely many disjoint open sets $\mathcal{O}_j$, $j = 1, 2, 3, \ldots$, containing parameters having the following properties:*

1. *If $\lambda \in \mathcal{O}_j$, then the Julia set of F_λ is a Sierpinski curve, so that if $\lambda \in \mathcal{O}_j$ and $\mu \in \mathcal{O}_k$, the Julia sets of F_λ and F_μ are homeomorphic;*
2. *But if $k \neq j$, the maps F_λ and F_μ are not topologically conjugate on their respective Julia sets.*

In Fig. 5 we display the parameter plane for the case $n = 2$ together with a magnification. The large central region is not a McMullen domain; rather it is a Sierpinski hole and it does not contain the origin.

We sketch the proof of this. We shall show that there are infinitely many open intervals in $\mathbb{R}^-$ in any neighborhood of the origin in parameter space in which the critical orbit eventually escapes. These Sierpinski holes need not lie along $\mathbb{R}^-$; we choose this just to simplify the proof.

Recall that, when $n = 2$, the four critical points and four prepoles of F_λ all lie on the critical circle of radius $|\lambda|^{1/4}$ centered at the origin. Also, the second image of all of the critical points is given by

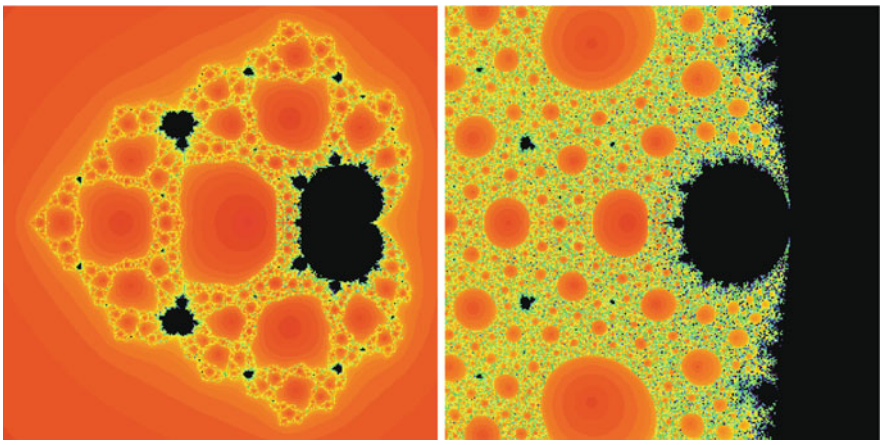


Fig. 5 The parameter plane for the family $z^2 + \lambda/z^2$ and a magnification centered at the origin

$$F_\lambda^2(c_\lambda) = 4\lambda + \frac{1}{4}$$

and so $\lambda \mapsto F_\lambda^2(c_\lambda)$ is an analytic function of λ that is a homeomorphism. If $-1/16 < \lambda < 0$, then one checks easily that the critical circle is mapped strictly inside itself. It follows that $J(F_\lambda)$ is a connected set and B_λ and T_λ are disjoint. In particular, the second image of the critical point lands on the real axis and lies in the complement of B_λ in $\mathbb{R}$.

We claim that there is an increasing sequence $\lambda_2,\lambda_3,\ldots$ in $\mathbb{R}^-$ with $\lambda_j \to 0$ and $F_{\lambda_j}^j(c_{\lambda_j}) = 0$ but $F_{\lambda_j}^i(c_{\lambda_j}) > 0$ for all $i < j$. To see this, note that, since $F_\lambda^2(c_\lambda) = 4\lambda + 1/4$, this quantity increases monotonically toward $1/4$ as $\lambda \to 0^-$. Now the orbit of $1/4$ remains in $\mathbb{R}^+$ for all iterations of F_0 and decreases monotonically to 0. Hence, given N, for λ sufficiently small, $F_\lambda^j(c_\lambda)$ also lies in $\mathbb{R}^+$ for $2 \le j \le N$ and moreover this finite sequence is decreasing. Now suppose $\beta < \alpha < 0$. We have $F_\beta(x) < F_\alpha(x)$ for all $x \in \mathbb{R}^+$. Also, $F_\beta^2(c_\beta) < F_\alpha^2(c_\alpha) < 1/4$. Hence $F_\beta^j(c_\beta) < F_\alpha^j(c_\alpha)$ for all j for which $F_\beta^j(c_\beta) \in \mathbb{R}^+$. The result then follows by continuity of F_λ with respect to λ.

Thus we have infinitely many Sierpinski holes in the parameter plane converging to 0 in $\mathbb{R}^-$. Since the escape times of these Sierpinski holes are all different, it follows that any two parameters drawn from diferent holes in this collection have non-conjugate dynamics (as shown in Section 5).

Note that $\lambda_2 = -1/16$. Using the previous observation, we may find open intervals I_j about λ_j for $j = 2, 3, \ldots$ having the property that, if $\lambda \in I_j$, then $F_\lambda^j(c_\lambda) \in T_\lambda$, and so $F_\lambda^{j+1}(c_\lambda) \in B_\lambda$. Therefore, $F_\lambda^n(c_\lambda) \to \infty$ as $n \to \infty$, and the Escape Trichotomy then shows that $J(F_\lambda)$ is a Sierpinski curve.

9 Julia Sets Converging to the Unit Disk

There is another way that the case $n = 2$ differs sharply from the case $n > 2$. Here the situation involves the structure of the Julia sets of F_λ when λ is close to 0. As we saw above, when $n > 2$ and $|\lambda|$ is small enough, λ lies in the McMullen domain and so the corresponding Julia sets are always Cantor sets of concentric simple closed curves. But when $n = 2$, the Julia sets vary wildly; often, but not always, they are Sierpinski curves. For example, in Fig. 5, note that there is a copy of the Mandelbrot set whose "tail" actually extends to the origin. Whenever λ is chosen in this set, $J(F_\lambda)$ contains small pieces that are homeomorphic to the corresponding Julia set from the quadratic family $z^2 + c$ together with infinitely many preimages of ∂B_λ (as well as other buried points).

In Fig. 6, note that as the parameter approaches 0, the size of the preimages of the trap door seems to decrease and the Julia set seems to grow "larger." This is indeed the case, for in [13] it was shown:

Theorem 7 *When $n = 2$, as $\lambda \to 0$, $J(F_\lambda)$ converges to the closed unit disk in the Hausdorff topology. On the other hand, when $n > 2$, this is not the case, as the*

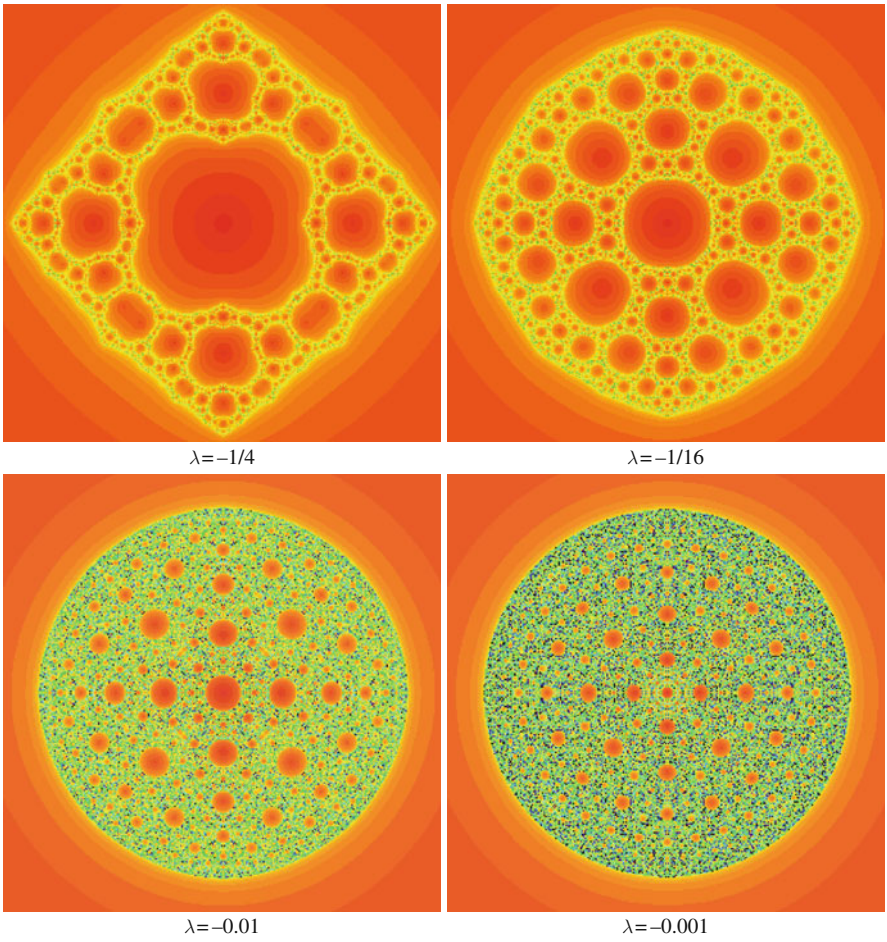


Fig. 6 Sierpinski curve Julia sets for various negative values of λ in the case $n = 2$. All of these sets are homeomorphic, but the dynamical behavior on each is very different

Fatou set always contains an annular component that contains a round annulus of some fixed width for all λ lying in some disk about the origin.

This first part of this result is somewhat surprising, since it is well known that Julia sets can never contain open sets unless the Julia set is the entire Riemann sphere. So here we find Julia sets getting arbitrarily close to the unit disk as $\lambda \to 0$. Of course, when $\lambda = 0$, there is an "implosion" and the Julia set is equal to the unit circle, not the unit disk.

The reason why these Julia sets converge to the closed unit disk $\mathbb{D}$ is as follows. Suppose that this is not the case. Then, given any sufficiently small $\epsilon > 0$, we may find a sequence of parameters $\lambda_j \to 0$ and another sequence of points $z_j \in \mathbb{D}$ such that $J(F_{\lambda_j}) \cap B_{2\epsilon}(z_j) = \emptyset$ for each j. Here $B_{2\epsilon}(z_j)$ is the disk of radius 2ϵ about z_j.

Since $\mathbb{D}$ is compact, there is a subsequence of the z_j that converges to some nonzero point $z^* \in \mathbb{D}$. For infinitely many parameters in the corresponding subsequence, we then have $J(F_{\lambda_j}) \cap B_\epsilon(z^*) = \emptyset$. Hence we may assume at the outset that we are dealing with a subsequence $\lambda_j \to 0$ such that $J(F_{\lambda_j}) \cap B_\epsilon(z^*) = \emptyset$.

Now consider the circle of radius $|z^*|$ centered at the origin. This circle meets $B_\epsilon(z^*)$ in an arc γ of length ℓ. Choose k so that $2^k \ell \geq 2\pi$. Since $\lambda_j \to 0$, we may choose j large enough so that $\left| F^i_{\lambda_j}(z) - z^{2^i} \right|$ is very small for $1 \leq i \leq k$, provided z lies outside the circle of radius $|z^*|/2$ centered at the origin. In particular, it follows that $F^k_{\lambda_j}(\gamma)$ is a curve whose argument increases by at least 2π, i.e., the curve $F^k_{\lambda_j}(\gamma)$ wraps at least once around the origin. As a consequence, the curve $F^k_{\lambda_j}(\gamma)$ must meet the Cantor necklace in the dynamical plane. But the boundary of this necklace lies in $J(F_{\lambda_j})$. Hence $J(F_{\lambda_j})$ must intersect this boundary. Since the Julia set is backward invariant, it follows that $J(F_{\lambda_j})$ must intersect $B_\epsilon(z^*)$. This then yields a contradiction, and so the result follows.

References

1. R. L. Devaney, D. M. Look, and D. Uminsky. The escape trichotomy for singularly perturbed rational maps. *Indiana Univ. Math. J.*, **54**, 267–285 (2005).
2. G. T. Whyburn. Topological characterization of the Sierpinski curve. *Fund. Math.*, **45**, 320–324 (1958).
3. C. McMullen. Automorphisms of rational maps. In *Holomorphic Functions and Moduli*. Vol. 1. Math. Sci. Res. Inst. Publ. 10. Springer, New York (1988).
4. J. Milnor. *Dynamics in One Complex Variable*. Princeton University Press, Princeton, NJ (2006).
5. R. L. Devaney. The McMullen domain: satellite Mandelbrot sets and Sierpinski holes. *Conform. Geom. Dynam.*, **11**, 164–190 (2007).
6. P. Roesch. On capture zones for the family f_(z) = z2 + λ/z2. In *Dynamics on the Riemann Sphere*. European Mathematical Society, Zurich, 121–130 (2006).
7. R. L. Devaney and K. M. Pilgrim. Dynamic classification of escape time Sierpinski curve Julia sets. To appear.
8. A. Douady and J. Hubbard. A proof of thurston's topological characterization of rational functions. *Acta. Math.*, **171**, 263–297 (1993).
9. R. L. Devaney. Structure of the McMullen domain in the parameter planes for rational maps. *Fundmen. Math.*, **185**, 267–285 (2005).
10. R. L. Devaney and S. Marotta. The McMullen domain: rings around the boundary. *Trans. AMS*, **359**, 3251–3273 (2007).
11. R. L. Devaney and D. M. Look. Buried Sierpinski curve Julia sets. *Discrete Contin. Dyn. Syst.*, **13**, 1035–1046 (2005).
12. P. Blanchard, R. L. Devaney, D. M. Look, P. Seal, and Y. Shapiro. Sierpinski curve julia sets and singular perturbations of complex polynomials. *Ergod. Theor. Dynamical Syst.*, **25**, 1047–1055 (2005).
13. R. L. Devaney and A. Garijo. Julia sets converging to the unit disk. *Proc. AMS*, **136**, 981–988 (2008).

Heteroclinic Switching in Coupled Oscillator Networks: Dynamics on Odd Graphs

Peter Ashwin, Gábor Orosz, and Jon Borresen

Abstract We review some examples of dynamics displaying sequential switching for systems of coupled phase oscillators. As an illustration we discuss a simple family of coupled phase oscillators for which one can find robust heteroclinic networks between unstable cluster states. For $N = 2k + 1$ oscillators we show that there can be open regions in parameter space where the heteroclinic networks have the structure of an odd graph of order k; a class of graphs known from permutation theory. These networks lead to slow sequential switching between cluster states that is driven by noise and/or imperfections in the system. The dynamics observed is of relevance to modelling the emergent complex dynamical behaviour of coupled oscillator systems, e.g. for coupled chemical oscillators and neural networks.

1 Introduction

Coupled phase oscillator networks provide a set of models that are very useful in a variety of applications ranging from theoretical and computational neuroscience [10, 11, 17, 18] to coupled chemical reactors [21]. They provide models that are amenable to variety of approaches aimed at understanding the emergent phenomena of such nonlinear dynamical systems. These approaches include "continuum approximations" as well as detailed studies of the dynamics and bifurcations of small numbers of oscillators; see [31] for a review. This paper reviews some recent results on switching dynamics for small numbers of oscillators. It is also explained how one of the dynamical structures (heteroclinic networks with odd graph structures) can be observed on scaling up to large numbers of oscillators.

The phase oscillator approach can be used in applications where the dynamics of individual elements can be modelled as a limit cycle oscillator and that can be

P. Ashwin (✉)
Mathematics Research Institute, College of Engineering, Mathematics and Physical Sciences
University of Exeter, Exeter, EX4 4QF, UK
e-mail: P.Ashwin@exeter.ac.uk

M. Thiel et al. (eds.), *Nonlinear Dynamics and Chaos: Advances and Perspectives,*
Understanding Complex Systems, DOI 10.1007/978-3-642-04629-2_3,

characterised by a scalar phase variable. This is particularly the case if there is weak coupling between the oscillators, or more specifically if the timescale associated with the coupling is much longer than the timescale associated with relaxation onto the limit cycle [6, 10, 17]. Moreover, phase models are typically useful even far from the weak coupling limit, especially concerning their predictions for generic bifurcations and attractors.

Clearly, the structure of coupling between oscillators is critical to determine what sort of dynamics is possible on a network of coupled oscillators, and there is a vast literature looking at the topology of coupling and the influence this has on the network dynamics; see for instance [1] and references therein. For the particular problem of synchronization this has been investigated by many authors, allowing for a variety of effects such as time-dependent coupling [7, 25] or delay in coupling [19, 32]. For the main part of this paper we do not address such issues, but rather explore the question of generic but complicated dynamics that can appear even for very simple all-to-all (fully symmetric) coupling. Indeed, this emphasises that not only the network structure, but also the actual coupling function is vital in determining the emergent dynamics of the system.

In this paper we review some recent work on the detailed dynamics of globally coupled phase oscillators, concentrating on the appearance of robust heteroclinic network attractors. These attractors manifest themselves in the dynamics as a "persistent transient" of slow switching between a number of unstable states that is driven by noise and/or imperfections. Such dynamics were first discovered in [17] and further studies have been made in [22, 23], mostly working with large populations. In Sect. 2 we review the fundamental dynamical principles of coupled phase oscillator systems with permutation symmetry. In Sect. 3 we discuss a detailed study of bifurcations leading to sequential slow switching attractors and show that they are only possible for $N \geq 4$; see [4]. In Sect. 4 we present a new result arising from [8] showing that a particular combinatorial structure can appear in the dynamics for coupling as in [17], for open sets of system parameters and arbitrarily large N.

Consider N identical phase oscillators that are coupled identically to each other

$$\dot{\theta}_i = \omega + \frac{K}{N} \sum_{j=1}^{N} g(\theta_i - \theta_j)\,, \tag{1}$$

where the dot denotes differentiation with respect to time t and $\theta_i \in \mathbb{T} = [0, 2\pi)$, $i = 1 \ldots N$, i.e. the state space is the N-torus $\mathbb{T}^N$. We will use the vector notation $\theta = \mathrm{col}[\,\theta_1, \ldots, \theta_N\,]$. The function $g\colon [0, 2\pi) \to \mathbb{R}$ is a 2π-periodic coupling function that we will assume to be smooth. We include a coupling strength parameter K for convenience but will set this to $K = 1$ by rescaling time.

For a given coupling function g and number of oscillators N, the dynamics of (1) can include a wide range of behaviours including in-phase (synchronous) oscillations, antiphase oscillations, and indeed arbitrarily complex partially synchronized cluster states that are stable [29]. However, system (1) can give much more complicated dynamics than just stable clustering. If the cluster states are of

saddle type, robust heteroclinic connections may be found. This is because cluster states may have unstable manifolds contained within the stable manifold of another state; see, for example, [2, 17, 22]. This is not possible for the coupling function $g(\varphi) = -\sin(\varphi)$ of the Kuramoto model [24] or the coupling function $g(\varphi) = -\sin(\varphi + \alpha)$ of the Kuramoto-Sakaguchi model [3], but it is possible for slightly more complicated coupling functions.

Choosing the coupling function

$$g(\varphi) = -\sin(\varphi + \alpha) + r\sin(2\varphi), \tag{2}$$

robust heteroclinic connections can be found that exist for open sets of parameter values on the (α, r) parameter plane [17]. Note that this coupling can be obtained from phase reduction of coupled Hodgkin-Huxley neurons with synaptic coupling [18]. More general couplings are considered for example in [12, 29]. We remark that system (1) can have highly nontrivial behaviour even for the Kuramoto coupling function if the frequencies are made non-identical; see [33]. In Sects. 3 and 4 we study the dynamics of certain parameter regimes for coupling (2), to illuminate the dynamics of system (1) and to better understand the graph structure of a possible heteroclinic network as N increases.

2 Dynamics and Bifurcations with S_N Symmetry

Here we review techniques from dynamical systems and bifurcation theory with symmetry [16, 15, 26, 27], noting that (1) is symmetric under all permutations of coordinates. In other words the dynamics is equivariant under the set S_N of permutations acting by

$$\sigma(\theta_1, \cdots, \theta_N) = (\theta_{\sigma(1)}, \cdots, \theta_{\sigma(N)}). \tag{3}$$

In addition to this, the system (1) has a symmetry given by translation of all components by the same phase shift: for all $\rho \in [0, 2\pi)$ we have invariance under

$$\rho(\theta_1, \cdots, \theta_N) = (\theta_1 + \rho, \cdots, \theta_N + \rho), \tag{4}$$

meaning that the system has symmetry $S_N \times S^1$. This has many dynamical consequences, the most obvious of which is that there will be an in-phase or fully synchronized solution

$$\theta_i(t) = \Omega t + r, \tag{5}$$

for $i = 1, \cdots, N$ where $\Omega = \omega + g(0)$ and r is constant. This solution may or may not be an attractor depending on its linear stability.

Another consequence of symmetry is that there will be a number of invariant subspaces forced by the symmetry; the most general of these being the "rotating

block" structures discussed in [6, 9]. We will focus on a particular case here; the M cluster states, for $1 \leq M \leq N$.

We define an M-cluster partition $\mathcal{A} = \{A_1, \ldots, A_M\}$ of $\{1, \ldots, N\}$ to be a partition such that

$$\{1, \ldots, N\} = \bigcup_{p=1}^{M} A_p \,, \tag{6}$$

where the A_p are pairwise disjoint sets ($A_p \cap A_q = \emptyset$ if $p \neq q$). Summing up the cluster sizes $a_p = |A_p|$ gives

$$\sum_{p=1}^{M} a_p = N \,. \tag{7}$$

Denoting the phase of the p-th cluster by ψ_p, i.e. defining $\psi_p := \theta_i = \theta_j = \theta_k = \ldots$ such that $\{i, j, k, \ldots\} \subset A_p$ we obtain

$$\dot{\psi}_p = \omega + \frac{1}{N} \sum_{q=1}^{M} a_q \, g(\psi_p - \psi_q) \,, \tag{8}$$

for $p = 1, \ldots, M$. The dynamics in such a subspace can be very complex, but initially we investigate simple clustering behaviour

$$\psi_p = \Omega \, t + \phi_p \,, \tag{9}$$

for $p = 1, \ldots, M$, where $\Omega \in \mathbb{R}^+$ and $\phi_p \in \mathbb{T}$. Since θ describes the phases of oscillators, (9) describes a periodic orbit in the state space of those oscillators.

Note that typical points in such a cluster state have symmetries obtained by all permutations within a cluster, i.e. an M-cluster state with partition $\mathcal{A}$ will have in terms of spatial symmetries

$$S_{a_1} \times \cdots \times S_{a_M} \,. \tag{10}$$

If there are any spatiotemporal symmetries these must be of the form of some cyclic permutation of clusters of equal size: they must be a semidirect product of permutations with the cycling

$$\left(S_{a_1} \times \cdots \times S_{a_\ell}\right)^k \otimes \mathbb{Z}_k \,, \tag{11}$$

where the $M = k\ell$ clusters permuted in groups of ℓ and $k\,(a_1 + \cdots + a_\ell) = N$ [6]. Note that the S^1 action is such that one can reduce the full dynamics of (1) to the dynamics of the phase differences and a translation on the S^1 orbit (group orbit). This means that, for example, the periodic orbits of (1) can be found by examining equilibria of the phase differences.

Linearisations of systems near equilibria with symmetries have structure forced upon them by the presence of symmetry; they must be block-diagonal with respect to the isotypic decomposition of the tangent space with respect to the group that fixes the equilibrium (the isotropy subgroup of the equilibrium). As a consequence, generic bifurcations with any symmetry group will have a centre manifold on which the action of the isotropy subgroup is an irreducible representation (irrep) [15, 16]. The permutation action of S_N on $\mathbb{T}^N$ neatly splits into two irreps, one which is a 1-dimensional trivial action and one which is an irreducible $(N-1)$-dimensional action. (There are other nontrivial irreps for S_N but they do not appear in (1).) Moreover, the nontrivial irrep corresponds precisely to the phase difference coordinates.

The generic S_N symmetry breaking bifurcations for the nontrivial irrep have been studied and classified; see, for example, [15]. However, the topology of the torus $\mathbb{T}^N$ can and does easily associate local symmetry breaking bifurcations with global bifurcations, as we will see in Sect. 3; see also [3, 6]. Typically, S_N symmetry breaking bifurcation from a stable fully symmetric state will result in a system with no nearby stable branches after bifurcation; this is associated in our system with global bifurcations that may lead to a stable cluster state or (only for $N \geq 4$) a robust heteroclinic attractor.

Note that periodic solutions of (1) with symmetry (11) can be interpreted in terms of cluster states in state space; for instance,

- S_N corresponds to in-phase solutions $\theta_k = \Omega t + \gamma$, $k = 1, \ldots, N$
- $\mathbb{Z}_N$ corresponds to anti-phase or "splay phase" solutions $\theta_k = \Omega t + \gamma + 2k\pi/N$, $k = 1, \ldots, N$
- $(S_{N/2})^2 \otimes \mathbb{Z}_2$ corresponds to antiphase solutions such as $\theta_k = \Omega t + \gamma$, $\theta_{k+N/2} = \Omega t + \gamma + \pi$, $k = 1, \ldots, N/2$

where $\Omega \in \mathrm{IR}+$ and $\gamma \in \mathbb{T}_j$ see (5) and (9).

Finally, we note that symmetry means that cluster states will appear in conjugate families; if a particular clustering is realised as a periodic orbit then so will all permutations of that clustering. This means that subspaces with conjugate isotropy subgroups will support identical dynamics.

3 Bifurcations for Three and Four Globally Coupled Oscillators

Returning to the specific system (1, 2), one can calculate the bifurcation behaviour in detail for small numbers of oscillators; this was done for $N = 3$ and $N = 4$ in [3] and here we briefly summarise some of the results, in particular with the aim of identifying where heteroclinic networks appear in parameter and state space.

Figure 1 shows the bifurcation diagram for $N = 3$ in the (r, α) plane and Table 1 describes briefly each of the lines of codimension-one bifurcations (both reproduced from [3]). The bifurcation diagram is shown surrounded by phase portraits in the $(\theta_1 - \theta_2, \theta_1 - \theta_3) \in [0, 2\pi)^2$ plane.

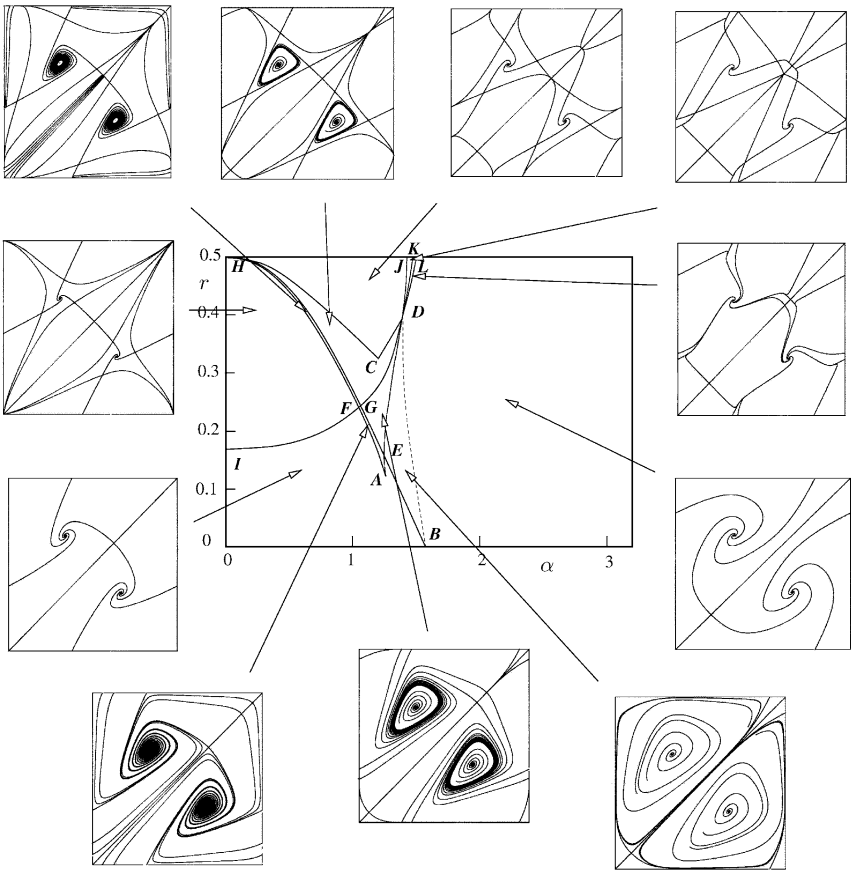


Fig. 1 Bifurcation curves in the (r, α) parameter plane for system (1) with coupling (2) in case of $N = 3$ (reproduced from [3]). The phase portraits are shown for $(\theta_1 - \theta_2, \theta_1 - \theta_3) \in [0, 2\pi)^2$. The codimension-one bifurcations are listed in Table 1. Note that there are heteroclinic/homoclinic cycles only on the *lines* BE, ED and HCD for $N = 3$. There are codimension-two bifurcations: A – cusp point; E – interaction of transcritical homoclinic and saddle-node homoclinic; D – interaction of saddle-node heteroclinic and saddle connection heteroclinic; H – degenerate Hopf bifurcation of antiphase

Moving on to the case for $N = 4$ oscillators, there are 12 invariant regions corresponding to points on $\mathbb{T}^4$ that lift to the set

$$\{(\theta_1, \theta_2, \theta_3, \theta_4) : \theta_{\sigma(1)} \leq \theta_{\sigma(2)} \leq \theta_{\sigma(3)} \leq \theta_{\sigma(4)} \leq \theta_{\sigma(1)} + 2\pi\}. \tag{12}$$

for permutations $\sigma \in S_4$. Figure 2 illustrates one of these invariant tetrahedra plotted in the $(\theta_1 - \theta_2, \theta_1 - \theta_3, \theta_1 - \theta_4) \in [0, 2\pi)^3$ space.

In Fig. 3 a bifurcation diagram is shown for the case $N = 4$ in the (r, α) plane, while Table 2 lists the codimension-one bifurcations (both reproduced from [3]).

Table 1 The letters label the curves of codimension-one bifurcations for $N = 3$ illustrated in Fig. 1 (reproduced from [3])

ID	Pitchfork bifurcation on invariant lines
BEGH	Transcritical bifurcation at 0
HFAED	Saddle-node bifurcation on invariant lines
BE	Transcritical homoclinic bifurcation
ED	Saddle-node homoclinic bifurcation
BD	Saddle-node of limit cycles
HCD	Saddle connection heteroclinic bifurcation
DJ	Pitchfork bifurcation
DK	Saddle-node bifurcation
DL	Saddle-node bifurcation

Observe that there is a similar set of bifurcations as in Fig. 3. In particular, for r small and $-\pi/2 < \alpha < \pi/2$ we have stable in-phase solutions while for r small and $\pi/2 < \alpha < 3\pi/2$ we have stable antiphase solutions. For both $N = 3$ and $N = 4$ the point $(r, \alpha) = (0, \pi/2)$ acts as an "organizing centre" for a number of lines of bifurcations that coincide at this point. Note that the Kuramoto-Sakaguchi coupling function (Eq. (2) with $r = 0$) is highly degenerate at the point $\alpha = \pi/2$.

In the shaded region *BEDTLV* of Fig. 3, the dynamics consists of robust heteroclinic cycles between symmetrically related cluster states with $S_2 \times S_2$ symmetry.

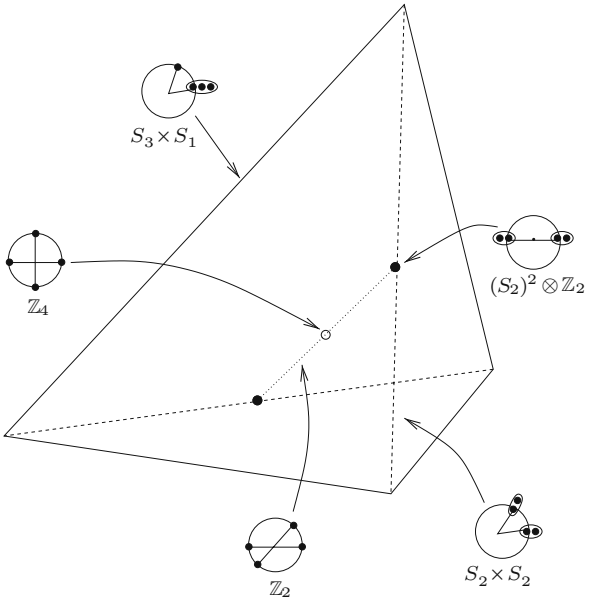


Fig. 2 Diagram showing one of the invariant regions (12) in state space for $N = 4$ oscillators (reproduced from [3]). This shows the relationship between the subspaces with differing symmetries. The point at the *centre* is the antiphase solution with $\mathbb{Z}_4$ symmetry and the faces of the invariant tetrahedron is made of points with S_2 symmetry

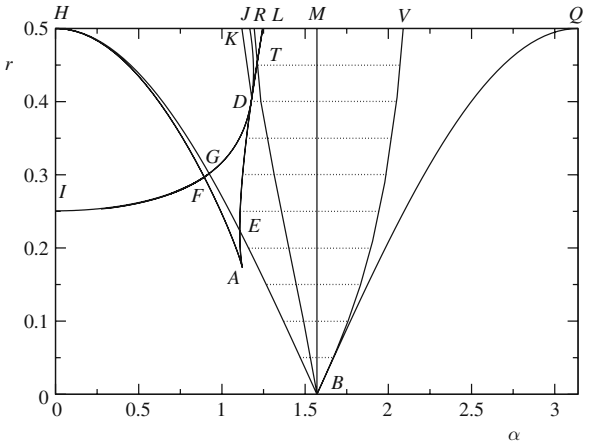


Fig. 3 Bifurcation diagram for $N = 4$ oscillators in the (α, r) plane (adapted from [3]); see Table 2 for a description of the bifurcation curves. There are robust heteroclinic cycles between 2-cluster states in the *shaded region outlined* by *BEDTLV*. These cycles are attractors to the *left* of the *line BM*. There is a complicated sequence of bifurcations in the near vicinity of point *D* that not is discussed here in detail

These cycles are attracting to the left of the curve *BM*. The $S_2 \times S_2$ symmetry corresponds to 2-cluster states where each cluster contains two oscillators. For example, the partition

$$\mathcal{A} = \{\{1,2\},\{3,4\}\}\,, \tag{13}$$

gives $[\,a_1, a_2\,] = [\,2,2\,]$ and the corresponding cluster phases can be defined as

$$\begin{aligned} \psi_1 &:= \theta_1 = \theta_2\,, \\ \psi_2 &:= \theta_3 = \theta_4\,. \end{aligned} \tag{14}$$

Table 2 A list of codimension-one bifurcations for $N = 4$ that are illustrated in Fig. 3 (reproduced from [3])

BEGH	Transcritical-pitchfork bifurcation at 0
BQ	Inverse pitchfork bifurcation of saddles at the point with $(S_2)^2 \otimes \mathbb{Z}_2$ symmetry
BV	Pitchfork/heteroclinic bifurcation of solutions with symmetry $S_2 \times S_2$ (in transversal to $S_2 \times S_2$ direction)
BM	Hopf bifurcation of antiphase points $\mathbb{Z}_4$ and change of stability of robust heteroclinic cycles
HAED	Saddle-node bifurcation to solutions with symmetry $S_3 \times S_1$
IFGD	Transcritical bifurcation of solutions with symmetry $S_3 \times S_1$
DK	Saddle connection bifurcation (not heteroclinic) in subspace with symmetry S_2
DJ	Transcritical bifurcation of solutions with symmetry $S_3 \times S_1$
DTL	Saddle-node bifurcation inside tetrahedra on S_2 plane
BTR	Pitchfork bifurcation of limit cycles within tetrahedron

Investigating the simple clustering behaviour (9) in these subspaces one may find the cluster states

$$[\,\psi_1,\psi_2\,] = \Omega\, t + [\,a,b\,]\,, \quad [\,\psi_1,\psi_2\,] = \Omega\, t + [\,b,a\,]\,, \tag{15}$$

where only the phase difference $a-b$ can be determined. By dropping Ωt we introduce the notation

$$P_1 = [\,a,a,b,b\,]\,, \quad P_2 = [\,b,b,a,a\,]\,, \tag{16}$$

for these cluster states.

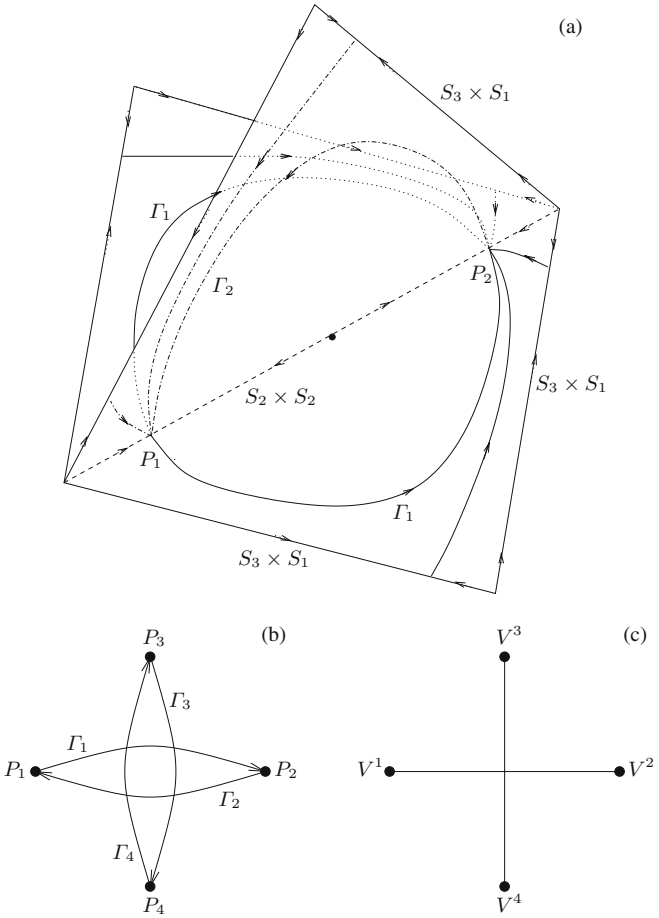


Fig. 4 Heteroclinic cycles for $N = 4$ oscillators. Panel (**a**) shows the detailed dynamics involving the saddle cluster states P_1, P_2 (16) (adapted from [3]). The cycle between states P_1, P_2 (16) and the cycle between states P_3, P_4 (17) are shown schematically in panel (**b**). Panel (**c**) represents the network as an undirected graph where the vertices $V^1, \ldots, V^4$ correspond to cluster states $P_1, \ldots, P_4$ and *edges* are drawn between them when heteroclinic connections exist. The parameters are from the shaded area *BEDTLV* in Fig. 3

One may find that the clusters with phase b are unstable to perturbations that split the oscillators and there exist heteroclinic orbits Γ_1 and Γ_2 that connect P_1 and P_2. Figure 4a illustrates this dynamics in state space (adapted from [3]). One may also find a similar cycle between the 2-cluster states

$$P_3 = [\, a, b, a, b \,], \quad P_4 = [\, b, a, b, a \,]. \tag{17}$$

corresponding to the partition

$$\tilde{\mathcal{A}} = \{\{1, 3\}, \{2, 4\}\}. \tag{18}$$

The heteroclinic cycles are sketched in Fig. 4b and the corresponding graph structure of the network is shown in Fig. 4c. The vertex V^m represents the cluster state P_m and vertices are connected by edges when there are heteroclinic connections between the corresponding cluster states. Observe that there exist switching dynamics between cluster states but the network structure is trivial: the heteroclinic network splits into two disconnected components. For $N \geq 5$ the network becomes connected and much more complex as will be shown in the next section.

4 Heteroclinic Networks for Odd Numbers of Oscillators

Here we consider some of the scaling properties of the robust heteroclinic networks discussed in the previous section, for more larger numbers of oscillators. As noted in [8] for odd numbers of oscillators, $N = 2k + 1$ ($k \geq 2$), one can obtain open regions of parameter space in which there are heteroclinic networks with the structure of an odd graph; in this paper we characterise this statement in detail and find the corresponding regions in parameter space numerically.

An *odd graph of order k* is a family of partition-defined graphs defined as follows. Let the vertex V^m represent a possible 2-cluster partition $\mathcal{A}^m = \left\{A_1^m, A_2^m\right\}$ of $N = 2k+1$ when $\left[\, a_1^m, a_2^m \,\right] = [\, k, k+1 \,]$. We call such partitions $[\, k, k+1 \,]$ cluster states; they clearly have $S_k \times S_{k+1}$ symmetry. There are $m = 1, \ldots, N!/(k!(k+1)!)$ such vertices. We say there is an edge from a vertex V^m to a vertex V^n if

$$A_1^n \subset A_2^m, \tag{19}$$

i.e. if the larger cluster of $\mathcal{A}^m$ contains the smaller cluster of $\mathcal{A}^n$. This graph is highly connected and each vertex has $k+1$ outgoing edges corresponding to which element in A_2^m that is not in A_1^n. The high level of connectivity of these graphs has resulted in suggestions to use the structure for fault-tolerant networks; see, for example, [14]. Examples of odd graphs of orders $k = 3$ and $k = 4$ are shown in Fig. 5.

We demonstrate that for system (1) with coupling (2) and any odd $N = 2k + 1$, $k = 2, 3, \ldots$ there is an open sets of parameters near $r = 0$ and $\alpha = \pi/2$ such that there is a heteroclinic network that contains the structure of an odd graph of

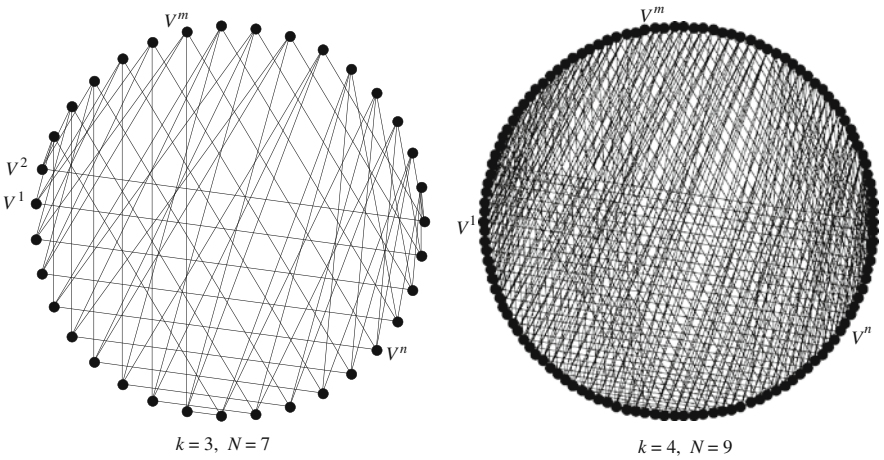


Fig. 5 Odd graphs of order $k = 3$ and $k = 4$; these structures are robustly present in heteroclinic networks of the coupled oscillator system (1) with coupling function (2) for $N = 7$ and $N = 9$ oscillators, respectively. The number of vertexes are 35 and 126, respectively, and they are presented in an analogous way to Fig. 4c

order k. This network is robust to perturbations of the coupling function g that are sufficiently small in the C^2 norm.

First, we show that two distinct families of $[\,k, k+1\,]$ cluster states may exist. We call these *ab* cluster states and *cd* cluster states, as will be clarified in Sect. 4.1; see also [2]. The *ab* cluster states are unstable to perturbations of the smaller cluster while the *cd* cluster states are unstable to perturbations of the larger cluster. Then we show that there may be heteroclinic connections from *ab* cluster states to *cd* cluster states and also from *cd* cluster states to *ab* cluster states. This means that applying perturbations to the appropriate (unstable) clusters one may navigate from a *ab* cluster state to another *ab* cluster state through a *cd* cluster state. Such navigation between the m-th and n-th *ab* cluster states is possible if and only if there is an edge between vertices V^m and V^n on the corresponding odd graph of order k; see Fig. 5. Notice that the odd graphs are not directed graphs: in fact if it is possible to navigate from the m-th *ab* cluster state to the n-th one, then it is possible to navigate the other way, through a different *cd* cluster state.

4.1 Existence, Stability and Connections of $[\,k, k+1\,]$ *Cluster States*

Consider the 2-cluster partition

$$\mathcal{A}^1 = \{\{1, \ldots, k\}, \{k+1, \ldots, 2k+1\}\} \tag{20}$$

so that $[\,a_1, a_2\,] = [\,k, k+1\,]$ (that is indeed a $[\,k, k+1\,]$ cluster state) and define the cluster phases as

$$\begin{aligned} \psi_1 &:= \theta_1 = \ldots = \theta_k\,, \\ \psi_2 &:= \theta_{k+1} = \ldots = \theta_{2k+1}\,. \end{aligned} \tag{21}$$

From (8), the equations for the time evolution of the cluster phases become

$$\begin{aligned} \dot{\psi}_1 &= \omega + \frac{1}{N}\big(a_1\, g(0) + a_2\, g(\psi_1 - \psi_2)\big)\,, \\ \dot{\psi}_2 &= \omega + \frac{1}{N}\big(a_1\, g(\psi_2 - \psi_1) + a_2\, g(0)\big)\,. \end{aligned} \tag{22}$$

Considering the periodic orbit (9) and introducing the notation $\phi := \phi_1 - \phi_2$ formula we have

$$0 = a_1(g(0) - g(-\phi)) + a_2(g(\phi) - g(0)) \tag{23}$$

to determine the phase difference ϕ.

Linearizing (22) one may obtain the tangential stability which has a trivial eigenvalue $\lambda_1 = 0$ and a tangential eigenvalue

$$\lambda_2 = \frac{1}{N}\big(a_1\, g'(-\phi) + a_2\, g'(\phi)\big)\,, \tag{24}$$

corresponding to perturbations that do not split either of the clusters. The linearization of (1) gives the transverse eigenvalues

$$\begin{aligned} \lambda_3 &= \frac{1}{N}\big(a_1\, g'(0) + a_2\, g'(\phi)\big)\,, \\ \lambda_4 &= \frac{1}{N}\big(a_1\, g'(-\phi) + a_2\, g'(0)\big)\,, \end{aligned} \tag{25}$$

where λ_3 corresponds to splitting of the cluster of k and has multiplicity $k-1$ while λ_4 corresponds to splitting of the cluster of $k+1$ and has multiplicity k.

Note that the dynamics in subspace (21) corresponding to partition $\mathcal{A}^1$ is effectively one-dimensional. Defining the variable $\psi := \psi_1 - \psi_2$ and subtracting the second equation from the first one in (22) results in

$$\dot{\psi} = \frac{1}{N}\big(a_1(g(0) - g(-\psi)) + a_2(g(\psi) - g(0))\big)\,. \tag{26}$$

There is an equilibrium $\psi(t) \equiv \phi$ determined by (23) while the related eigenvalue is given by (24). Substituting (2) and $N = 2k+1$ into (26) and considering the large k limit we have

$$\dot{\psi} = -\cos\alpha \sin\psi + r\sin(2\,\psi) + \mathcal{O}\Big(\frac{1}{2k+1}\Big)\,. \tag{27}$$

Now consider the 2-cluster partition

$$\mathcal{A}^2 = \{\{k+2,\,\ldots,2k+1\},\{1,\ldots,k+1\}\} \tag{28}$$

for which $[\,\tilde{a}_1,\tilde{a}_2\,] = [\,k,k+1\,]$ and the corresponding cluster phases can be defined as

$$\begin{aligned}\tilde{\psi}_1 &:= \theta_{k+2} = \ldots = \theta_{2k+1}\,,\\ \tilde{\psi}_2 &:= \theta_1 = \ldots = \theta_{k+1}\,.\end{aligned} \tag{29}$$

The evolution of these phases are still given by (22) so one may find the same equilibria with the same stability properties as for partition $\mathcal{A}^1$ (20).

According to (19), since $\{k+2,\,\ldots,2k+1\} \subset \{k+1,\ldots,2k+1\}$ then on the odd graph of order k there is an edge connecting the vertices V^1 and V^2 (that represent the partitions $\mathcal{A}^1$ (20) and $\mathcal{A}^2$ (28), respectively). We show that there is a heteroclinic connection from the *ab* cluster state with partition $\mathcal{A}^1$ to the *cd* cluster state with partition $\mathcal{A}^2$ and that there is a connection from the *cd* cluster state with partition $\mathcal{A}^2$ to the *ab* cluster state with the same partition. Thus one may navigate from the *ab* cluster state with partition $\mathcal{A}^1$ to the *ab* cluster state with partition $\mathcal{A}^2$ through the *cd* cluster state with partition $\mathcal{A}^2$.

Thus, we are interested in parameter regions where **A**, **B** and **C** hold:

A There exist two different cluster states in the subspace (21) such that

$$[\,\psi_1,\psi_2\,] = \Omega\,t + [\,a,b\,]\,,\quad [\,\psi_1,\psi_2\,] = \Omega\,t + [\,c,d\,]\,, \tag{30}$$

that are both $[\,k,k+1\,]$ cluster states with partition $\mathcal{A}^1$ (20). They are called *ab* and *cd* cluster states and by dropping the $\Omega\,t$ we loosely say that in a *ab* cluster state k oscillators have phase a and $k+1$ oscillators have phase b while in a *cd* cluster state k oscillators have phase c and $k+1$ oscillators have phase d. Indeed, only the phase differences

$$\phi^{ab} = a - b\,,\quad \phi^{cd} = c - d\,, \tag{31}$$

can be determined from (23) and by symmetry, such periodic orbits exist for all other $[\,k,k+1\,]$ cluster states with different partitions (e.g. with $\mathcal{A}^2$ (28)).

B We require that the *ab* cluster states are unstable to perturbations of the larger cluster of *b*-s and the *cd* cluster states are unstable to perturbation of the smaller cluster of *c*-s while all other perturbations decay. This can be characterised in saying that we wish to find α, r such that for *ab* cluster states we have $\lambda_2 < 0$, $\lambda_3 < 0$, $\lambda_4 > 0$ while for *cd* cluster states we have $\lambda_2 < 0$, $\lambda_3 > 0$, $\lambda_4 < 0$; see (24) and (25).

C We wish to find a subregion of **A, B** in parameter space such that there is a heteroclinic connection from the *ab* cluster state with partition $\mathcal{A}^1$ (20) to the *cd* cluster state with partition $\mathcal{A}^2$ (28) and a heteroclinic connection from the *cd* cluster state with partition $\mathcal{A}^2$ to the *ab* cluster state with partition $\mathcal{A}^2$.

Steps **A, B** above are routine in terms of root finding and then evaluating the stabilities but step **C** is more complex. Note that (similarly to the $N = 4$ case in Fig. 4a) there is an additional $[\,k, k+1\,]$ cluster state in the subspace (21) but that is unstable against tangential perturbations ($\lambda_2 > 0$) and hence does not appear in the heteroclinic network.

First let us search for the $ab \to cd$ heteroclinic connection by examining the partition

$$\mathcal{B}^1 = \{\{1, \ldots, k\}, \{k+1\}, \{k+2, \ldots, 2k+1\}\}\,, \tag{32}$$

that corresponds to (by abuse of notation) $[\,b_1, b_s, b_2\,] = [\,k, 1, k\,]$. Defining the cluster phases as

$$\begin{aligned} \psi_1 &:= \theta_1 = \ldots = \theta_k\,, \\ \psi_s &:= \theta_{k+1}\,, \\ \psi_2 &:= \theta_{k+2} = \ldots = \theta_{2k+1}\,, \end{aligned} \tag{33}$$

leads to

$$\begin{aligned} \dot{\psi}_1 &= \omega + \frac{1}{N}\big(b_1\, g(0) + b_s\, g(\psi_1 - \psi_s) + b_2\, g(\psi_1 - \psi_2)\big)\,, \\ \dot{\psi}_s &= \omega + \frac{1}{N}\big(b_1\, g(\psi_s - \psi_1) + b_s\, g(0) + b_2\, g(\psi_s - \psi_2)\big)\,, \\ \dot{\psi}_2 &= \omega + \frac{1}{N}\big(b_1\, g(\psi_2 - \psi_1) + b_s\, g(\psi_2 - \psi_s) + b_2\, g(0)\big)\,, \end{aligned} \tag{34}$$

see (8). Notice that for $\psi_s = \psi_2$ one obtains a $[k, k+1]$ cluster state with partition $\mathcal{A}^1$ (20) and (21), While for $\psi_s = \psi_1$ $a[k, k+1]$ cluster state with partition A^2 (28) and (29) is obtained. One can numerically verify that there exists a heteroclinic connection between the *ab* cluster state $[\,\psi_1, \psi_s, \psi_2\,] = \Omega\, t + [\,a, b, b\,]$ and the *cd* cluster state $[\,\psi_1, \psi_s, \psi_2\,] = \Omega\, t + [\,d, d, c\,]$. Figure 6 shows such connections for $k = 10$ ($N = 21$) with parameters $\alpha = 1.52$ and $r = 0.1$. To eliminate $\Omega\, t$ the phase differences $\psi_1 - \psi_s$ and $\psi_s - \psi_2$ are used.

Now let us search for the $cd \to ab$ connection by examining the partition

$$\mathcal{B}^2 = \{\{k+2, \ldots, 2k\}, \{2k+1\}, \{1, \ldots, k+1\}\}\,, \tag{35}$$

that corresponds to $[\,\tilde{b}_1, \tilde{b}_s, \tilde{b}_2\,] = [\,k-1, 1, k+1\,]$. Now defining the cluster phases as

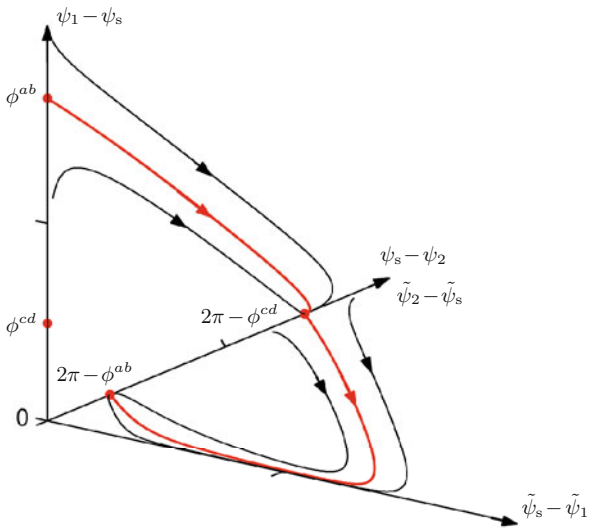


Fig. 6 Heteroclinic connections (*thick curves*) connecting *ab* and *cd* cluster states (*dots*) for $k = 10$ ($N = 21$). The figure shows that one may navigate from the *ab* cluster state with partition $\mathcal{A}^1$ (20) to the *ab* cluster state with partition $\mathcal{A}^2$ (28) through the *cd* cluster state with partition $\mathcal{A}^2$, that is, the sequence of connections $ab \to cd \to ab$ corresponds to an edge between the vertices V^1 and V^2 (representing $\mathcal{A}^1$ and $\mathcal{A}^2$) on the odd graph of order k, see Fig. 5. Additional trajectories (*thin curves*) demonstrate that the connections are attracting, that is, the heteroclinic connections are sinks within the relevant subspaces. The parameters are $\alpha = 1.52$ and $r = 0.1$ correspond to the cross in Fig. 7d

$$\begin{aligned} \tilde{\psi}_1 &:= \theta_{k+2} = \ldots = \theta_{2k}\,, \\ \tilde{\psi}_s &:= \theta_{2k+1}\,, \\ \tilde{\psi}_2 &:= \theta_1 = \ldots = \theta_{k+1}\,. \end{aligned} \tag{36}$$

leads to the same set of equations as (34) for $\tilde{\psi}_p$-s but with the current set of $\tilde{b}_p$-s. Note that for $\tilde{\psi}_s = \tilde{\psi}_1$ a $[\,k, k+1\,]$ cluster state is obtained with partition $\mathcal{A}^2$ (28) and (29).

We can numerically verify that there exist heteroclinic connections between the *cd* cluster state $[\,\tilde{\psi}_1, \tilde{\psi}_s, \tilde{\psi}_2\,] = \Omega\, t + [\,c, c, d\,]$ and the *ab* cluster state $[\,\tilde{\psi}_1, \tilde{\psi}_s, \tilde{\psi}_2\,] = \Omega\, t + [\,a, a, b\,]$ as shown in Fig. 6 for $k = 10$ ($N = 21$) with parameters $\alpha = 1.52$, $r = 0.1$. Now the phase differences $\tilde{\psi}_s - \tilde{\psi}_1$ and $\tilde{\psi}_2 - \tilde{\psi}_s$ are used to represent the connection. Note that in Fig. 6 only the plane spanned by $\psi_1 - \psi_s$ and $\psi_s - \psi_2$ and

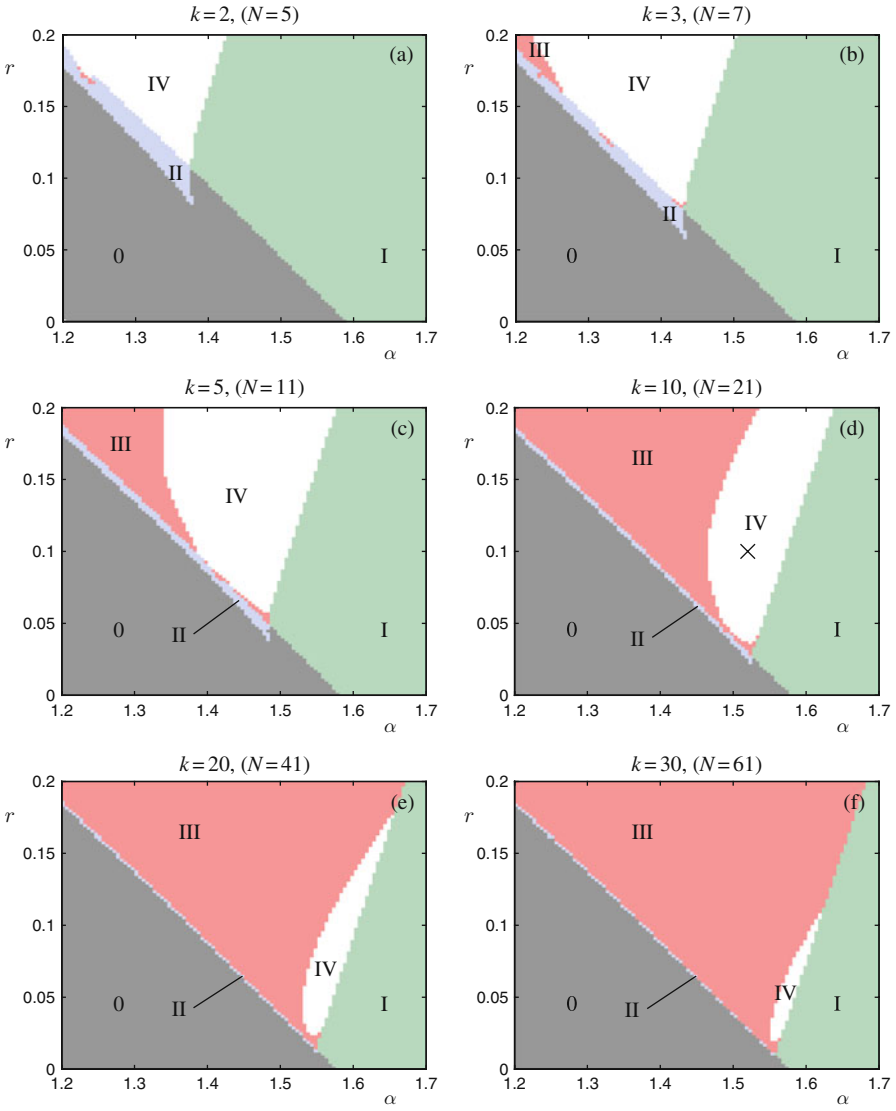


Fig. 7 In (α, r) parameter space the *white region* IV shows where the $N = 2k+1$ oscillator system $(1, 2)$ has a heteroclinic network between $[\, k, k+1\,]$ cluster states with the structure of the odd graph of order k for (**a**) $k = 2$, (**b**) $k = 3$, (**c**) $k = 5$, (**d**) $k = 10$, (**e**) $k = 20$ and (**f**) $k = 30$. The other regions indicated are as follows. *Dark grey region* 0 – neither cluster state exists; *light green region* I – only *ab* cluster states exist; and *light blue region* II – only *cd* cluster states exist; *red region* III – both type of cluster states exist but connections are missing. Note that the *white region* is open (confirming that the cycle is robust) but becomes smaller and moves towards $(\alpha, r) = (\pi/2, 0)$ on increasing k. The cross in panel (**d**) corresponds to the simulation in Fig. 6

the plane spanned by $\tilde{\psi}_s - \tilde{\psi}_1$ and $\tilde{\psi}_2 - \tilde{\psi}_s$ are used by the dynamics; the plane spanned by $\psi_1 - \psi_s$ and $\tilde{\psi}_s - \tilde{\psi}_1$ is transverse to the dynamics.

For values of parameters α, r and k where **A, B** hold one may run finite time simulations of system (34) and determine whether the above described $ab \to cd$ and $cd \to ab$ connections exist. In this way the existence of a heteroclinic network equivalent in structure to an odd graph of order k can be it verified numerically. This could in principle be done to a finite precision, provided that a bound on the errors is also calculated. In Fig. 7, a 100×100 mesh is used over the parameter regime $\alpha \in (1.2, 1.7), r \in (0, 0.2)$ and at each mesh-point it is checked whether the conditions **A, B** and **C** are satisfied for k ranging from $k = 2$ ($N = 5$) up to $k = 30$ ($N = 61$).

Region IV (white) shows where **A, B** and **C** hold while region III (red) shows where only **A, B** hold but not **C**. Regions I (light green) and II (light blue) are where **A, B** are only partially satisfied, i.e. only *ab* or *cd* cluster states exist, respectively. Finally, considering parameters from region 0 (dark grey) no condition hold, that is no cluster states can be found. Observe that the white region IV shrinks as k increases.

Although this is not a rigorous proof that the odd graph structure is present for arbitrary (finite) k, it does suggest a means of verifying this. It may be possible to analytically prove the existence of an open non-empty region in the (α, r) parameter plane satisfying **A, B** and **C** by using rigorous singular perturbation methods with a small parameter $\epsilon = 1/k$.

5 Discussion

We have shown that remarkably complicated bifurcations and structures may emerge in the dynamics of (1). In particular, robust heteroclinic networks can appear that lead to slow sequential switching between cluster states. We found that the number of oscillators needs to satisfy $N \geq 5$ to give rise to nontrivial robust switching dynamics between cluster states. The reason for this is that 5 is the minimum number where there is a nontrivial asymmetric partition, namely a $[2, 3]$ cluster state. Such a state has two clusters of differing size that can have different transverse stabilities, and this permits a nontrivial switching between different 2-cluster partitions. In general for $N = 2k + 1$ oscillators, we determined parameter regimes where $[k, k + 1]$ cluster states are connected to form a heteroclinic network with the structure of an odd graph of order k. We remark that for $N \geq 7$ the odd graph heteroclinic network is only a subnetwork of the full attracting heteroclinic network. The full network includes connections not just between cluster states of type $[k, k + 1]$ but also between cluster states of type $[k - 2m, k + 2m + 1]$ for a variety of m. Up to now we have only scratched the surface of the combinatorial complexity that may exist in such networks.

We note that it is possible to find rather different heteroclinic structures for a variety of different coupling functions. For example, considering

$$g(\varphi) = -\sin(\varphi + \alpha) + r\sin(2\varphi + \beta), \tag{37}$$

and $N = 2k + 1$ ($N \geq 5$), it is possible to find robust heteroclinic cycles between $[k, k, 1]$ cluster states for open sets of parameters [5, 28, 34]. We also remark that (1) does not have any sort of variational structure or global Lyapunov function for general choices of g (although it may do for specific cases of interest [11]). Lack of variational structure is clearly a necessary condition for the existence of heteroclinic connections in the dynamics.

On preserving the coupling structure but breaking the permutation symmetry by making the oscillators non-identical, the dynamics can still be efficiently explained with reference to the heteroclinic network [28, 34]. Moreover, for strong nonlinear coupling, one can explain extreme sensitivity to detuning of the frequencies of oscillators. This is where arbitrarily small differences in the frequencies give rise to loss of frequency locking and this depends apparently on the presence of robust heteroclinic attractors that are topologically non-trivial on $\mathbb{T}^N$ [4]. On breaking the symmetry of coupling structure by removing connections, the dynamics can be much more rich; see, for example [13, 30] and references within. However, heteroclinic connections between unstable states may still exist even when the coupling is not all-to-all leading to sequential switching [1, 20].

The switching dynamics occurring in networks provides some promising models for a variety of phenomena, especially for neural applications where synchrony tends to be frustrated due to the competing requirements of efficiency (a neural code needs a minimum of energy to be maintained) and robustness (a neural code should be insensitive to removal or malfunction of individual cells). Particular applications that may be of relevance to the models presented here include finite-state computation [2] and spatio-temporal code generation [28, 34].

Acknowledgement We gratefully acknowledge Oleksandr Burylko, Mike Field, Yuri Maistrenko, Jeff Moehlis, Marc Timme, Stuart Townley and John Wordsworth for discussions related to this work. We also thank Robin Chapman for originally pointing us to the literature on odd graphs. This research was supported by EPSRC under grant EP/C510771/1 and by the Institute for Collaborative Biotechnologies under grant DAAD19-03-D004 from the US Army Research Office.

References

1. M. Aguiar, P. Ashwin, A. Dias, and M. Field. Dynamics of coupled cell networks: synchrony, heteroclinic cycles and inflation. Preprint (2010).
2. P. Ashwin and J. Borresen. Encoding via conjugate symmetries of slow oscillations for globally coupled oscillators. *Phys. Rev. E*, **70**(2), 026203 (2004).
3. P. Ashwin, O. Burylko, and Y. Maistrenko. Bifurcation to heteroclinic cycles and sensitivity in three and four coupled phase oscillators. *Phys. D*, **237**(4), 454–466 (2008).
4. P. Ashwin, O. Burylko, Y. Maistrenko, and O. Popovych. Extreme sensitivity to detuning for globally coupled phase oscillators. *Phys. Rev. Lett.*, **96**(5), 054102 (2006).
5. P. Ashwin, G. Orosz, J. Wordsworth, and S. Townley. Dynamics on networks of cluster states for globally coupled phase oscillators. *SIAM J. Appl. Dynam. Syst.*, **6**(4), 728–758 (2007).

6. P. Ashwin and J. W. Swift. The dynamics of *n* weakly coupled identical oscillators. *J. Nonlinear Sci*., **2**(1), 69–108 (1992).
7. I. Belykh, V. Belykh, and M. Hasler. Generalized connection graph method for synchronization in asymmetrical networks. *Phys. D*, **224**(1–2), 42–51 (2006).
8. J. Borresen. *Dynamical encoding in systems of globally coupled oscillators*. PhD thesis, University of Exeter (2006).
9. E. Brown, P. Holmes, and J. Moehlis. Globally coupled oscillator networks. In E. Kaplan, J.E. Marsden, and K.R. Sreenivasan, editors, *Perspectives and Problems in Nonlinear Science: A Celebratory Volume in Honor of Larry Sirovich*. Springer, New York, pp. 183–215 (2003).
10. E. Brown, J. Moehlis, and P. Holmes. On the phase reduction and response dynamics of neural oscillator populations. *Neural Comput*., **16**(4), 673–715 (2004).
11. A. C. C. Coolen, R. Kühn, and P. Sollich. *Theory of Neural Information Processing Systems*. Oxford University Press, Oxford (2005).
12. H. Daido. Generic scaling at the onset of macroscopic mutual entrainment in limit-cycle oscillators with uniform all-to-all coupling. *Phys. Rev. Lett*., **73**(4), 760–763 (1994).
13. M. Field. Combinatorial dynamics. *Dynam. Syst*., **19**(3), 217–243 (2004).
14. A. Ghafoor and T. R. Bashkow. A study of odd graphs as fault-tolerant interconnection networks. *IEEE Trans. Comput*., **40**(2), 225–232 (1991).
15. M. Golubitsky and I. Stewart. *The Symmetry Perspective: From Equilibrium to Chaos in Phase Space and Physical Space, Progress in Mathematics*. Birkäuser, Berlin, vol. 200 (2002).
16. M. Golubitsky, I. Stewart, and D. G. Schaeffer. *Singularities and Groups in Bifurcation Theory: Volume II, Applied Mathematical Sciences*. Springer, New York, vol. 69 (1988).
17. D. Hansel, G. Mato, and C. Meunier. Clustering and slow switching in globally coupled phase oscillators. *Phys. Rev. E*, **48**(5), 3470–3477 (1993a).
18. D. Hansel, G. Mato, and C. Meunier. Phase dynamics for weakly coupled Hodgkin-Huxley neurons. *Europhys. Lett*., **23**(5), 367–372 (1993b).
19. E. M. Izhikevich. Phase models with explicit time delays. *Phys.l Rev. E*, **58**(1), 905–908 (1998).
20. Ö. Karabacak and P. Ashwin. Heteroclinic ratchets in networks of coupled oscillators. *J. Nonlinear sci.*, **20**(1), 105–129 (2010).
21. I. Z. Kiss, C. G. Rusin, H. Kori, and J. L. Hudson. Engineering complex dynamical structures: sequential patterns and desynchronization. *Science*, **316**(5833), 1886–1889 (2007).
22. H. Kori. Slow switching in a population of delayed pulse-coupled oscillators. *Phys. Rev. E*, **68**(2), 021919 (2003).
23. H. Kori and Y. Kuramoto. Slow switching in globally coupled oscillators: robustness and occurrence through delayed coupling. *Phys. Rev. E*, **63**(4), 046214 (2001).
24. Y. Kuramoto. Self-entrainment of a population of coupled non-linear oscillators. In H. Araki, editor, *International Symposium on Mathematical Problems in Theoretical Physics, Lecture Notes in Physics*. Springer, New York, vol. 69, pp. 420–422 (1975).
25. W. L. Lu, F. M. Atay, and J. Jost. Chaos synchronization in networks of coupled maps with time-varying topologies. *Eur. Phys. J. B*, **63**(3), 399–406 (2008).
26. J. Moehlis and E. Knobloch. Equivariant bifurcation theory. *Scholarpedia*, **2**(9), 2511 (2007a).
27. J. Moehlis and E. Knobloch. Equivariant dynamical systems. *Scholarpedia*, **2**(10), 2510 (2007b).
28. G. Orosz, P. Ashwin, and S. Townley. Learning of spatio-temporal codes in a coupled oscillator system. *IEEE Trans. Neural Netw*., 20(7), 1135–1147 (2009).
29. G. Orosz, J. Moehlis, and P. Ashwin. Designing the dynamics of globally coupled oscillators. *Prog. Theor. phys.*,122(3), 611–630 (2009).
30. I. Stewart and M. Parker. Periodic dynamics of coupled cell networks II: cyclic symmetry. *Dynam. Syst*., **23**(1), 17–41 (2008).
31. S. H. Strogatz. From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. *Phys. D*, **143**(1–4), 1–20 (2000).
32. M. Timme, F. Wolf, and T. Geisel. Prevalence of unstable attractors in networks of pulse-coupled oscillators. *Phys. Rev. Lett*., **89**(15), 154105 (2002).

33. L. S. Tsimring, N. F. Rulkov, M. L. Larsen, and M. Gabbay. Repulsive synchronization in an array of phase oscillators. *Phys. Rev. Lett*., **95**(11), 014101 (2005).
34. J. Wordsworth and P. Ashwin. Spatio-temporal coding of inputs for a system of globally coupled phase oscillators. *Phys. Rev. E*, **78**(6), 066203 (2008).

Dynamics of Finite-Size Particles in Chaotic Fluid Flows

Julyan H.E. Cartwright, Ulrike Feudel, György Károlyi, Alessandro de Moura, Oreste Piro, and Tamás Tél

Abstract We review recent advances on the dynamics of finite-size particles advected by chaotic fluid flows, focusing on the phenomena caused by the inertia of finite-size particles which have no counterpart in traditionally studied passive tracers. Particle inertia enlarges the phase space and makes the advection dynamics much richer than the passive tracer dynamics, because particles' trajectories can diverge from the trajectories of fluid parcels. We cover both confined and open flow regimes, and we also discuss the dynamics of interacting particles, which can undergo fragmentation and coagulation.

1 Introduction and Overview

A correct formulation of the problem of the motion of finite-size particles in fluid flows has presented difficult challenges for generations of fluid dynamicists. Although in principle this problem is "just" another application of the Navier–Stokes equation, with moving boundary conditions, a direct solution of the fluid dynamical equations is not only very difficult, but also not very illuminating. So from the nineteenth century onwards efforts were made to find the appropriate approximations which allow one to write the equations of motion of small rigid particles in a given flow in the form of ordinary differential equations, regarding the flow's velocity field as given. Some very subtle issues are involved in making the right kinds of approximations and assumptions in a self-consistent way, and a number of incorrect results appeared in the early literature. The issue was finally resolved when Maxey and Riley [1] wrote down the equations of motion for a small spherical rigid particle advected by a (smooth) flow and Auton et al. [2], following Taylor's work [3], corrected the form of the added-mass term.

A. de Moura (✉)
Institute of Complex Systems and Mathematical Biology, University of Aberdeen, King's College, Meston Building, AB24 3UE Aberdeen, United Kingdom
e-mail: a.moura@abdn.ac.uk

M. Thiel et al. (eds.), *Nonlinear Dynamics and Chaos: Advances and Perspectives,* Understanding Complex Systems, DOI 10.1007/978-3-642-04629-2_4,

The Maxey-Riley equations allow the global dynamics of a single advected finite-size particle to be investigated with the techniques of dynamical systems theory. The dynamics of advected particles in fluid flows have been a favourite subject of investigation of chaos and related complex dynamical regimes since the pioneering work of Aref [4], but all the early works assumed that the particles' size and their inertia could be neglected – the *passive tracer* assumption. Finite size results in inertia, which introduces a new richness to the dynamics, since finite-size particles are no longer enslaved to the motion of the flow surrounding them – they have their own dynamics, distinct from that of the fluid. A whole new world of challenges and possibilities opens up to dynamicists once finite size and inertia are considered. This gives this subject a great importance from the theoretical point of view alone; not to mention its practical importance: polluting particles suspended in the atmosphere and plankton organisms floating in the ocean are just a few of the systems whose understanding involves the theory of the dynamics of finite-size particles in complex flows.

This chapter presents an overview of the subject of the dynamics of finite-size particles in chaotic flows, focusing on a few chosen topics of current research. The choice of topics reflects the authors' own research interests; we make no apologies for that: we in no way claim this to be an exhaustive review on this area. But we do think the topics we cover here give the reader a good idea of what is going on in this exciting area of research.

We first introduce the Maxey-Riley equation in Sect. 2. Its assumptions and range of validity are discussed, as well as some of its basic consequences; but we do not show the derivations. The dynamics of finite-size particles can often be understood by using the simpler dynamics of passive tracers as a starting point. In Sect. 3 we discuss the chaotic advection of non-inertial tracers, and also introduce some of the flows which will be used as examples in later sections.

The dynamics of non-interacting finite-size particles is the subject of Sects. 4 and 5. Section 4 deals with confined flows, whereas Sect. 5 focuses on open flows. These two kinds of flow have very different long-time behaviours, which lead to quite distinct particle dynamics. In these sections, we focus especially on the new phenomena caused by the particles' inertia, which are not present in the case of passive tracers. The challenging and very important topic of interacting finite-size particles is covered by Sect. 6, which reviews recent results on the processes of fragmentation and coagulation of finite-size particles. Finally, in Sect. 7, we make some remarks on the future directions of this research area.

2 Motion of Finite-Size Particles in Fluid Flows

When studying the motion of particles advected by fluid flows, it is commonly assumed that one can consider the particles as *passive tracers*, with negligible mass and size. This amounts to neglecting the particles' inertia, and assuming that they take on the velocity of the surrounding fluid, instantaneously adapting to any changes in the fluid velocity. If $\mathbf{u}$ is the (possibly time-dependent) fluid's velocity

field, and denoting by $\mathbf{r}(t)$ the position of a particle, the passive tracer assumption implies that $\mathbf{r}$ satisfies the differential equation:

$$\dot{\mathbf{r}}(t) = \mathbf{u}(\mathbf{r}(t),t). \tag{1}$$

The passive tracer assumption is extensively used in fluid dynamics [5, 6], and it is a good approximation in a number of cases. There are many situations, however, where it does not apply, and we need to take into account the fact that particles have finite sizes and masses (for reviews see [7–9]). Finite-size particles are not able to adjust their velocities instantaneously to that of the fluid, and in addition their density may be different from that of the fluid. Therefore, in general the particle velocity differs from the fluid velocity. This means that the dynamics of finite-size particles is far richer and more complex than that of passive tracers.

2.1 The Maxey-Riley Equation

In order to study the dynamics of finite-size particles advected by chaotic flows, we need to have a simple formulation of the equations of motion of the advected particles. The problem is that finite-size particles are actually extended objects with their own boundaries. The rigorous way to analyse their dynamics would involve solving the Navier–Stokes equation for moving boundaries, with all the complications this implies. The partial differential equations resulting from this approach would be very difficult to solve and analyse; and as dynamicists, we would like to have the particle's motion described by ordinary differential equations, similar to Eq. (1). Fortunately, an approximate differential equation for the motion of small spherical particles in flows may be written down [1, 2]. If a particle has radius a and mass m_p, its motion is given to a good approximation by the *Maxey-Riley equation*:

$$\begin{aligned} m_p\dot{\mathbf{v}} = m_f\frac{D}{Dt}\mathbf{u}(\mathbf{r}(t),t) - \frac{1}{2}m_f\left(\dot{\mathbf{v}} - \frac{D}{Dt}\left[\mathbf{u}(\mathbf{r}(t),t) + \frac{1}{10}a^2\nabla^2\mathbf{u}(\mathbf{r}(t),t)\right]\right) \\ -6\pi a\rho_f\nu\mathbf{q}(t) + (m_p - m_f)\mathbf{g} - 6\pi a^2\rho_f\nu\int_0^t d\tau\,\frac{d\mathbf{q}(\tau)/d\tau}{\sqrt{\pi\nu(t-\tau)}}, \end{aligned} \tag{2}$$

where

$$\mathbf{q}(t) \equiv \mathbf{v}(t) - \mathbf{u}(\mathbf{r}(t),t) - \frac{1}{6}a^2\nabla^2\mathbf{u}.$$

Here $\mathbf{r}(t)$ and $\mathbf{v}(t) \equiv d\mathbf{r}(t)/dt$ are the position and velocity of the particle, respectively, and $\mathbf{u}(\mathbf{r},t)$ is the undisturbed flow field at the location of the particle. m_f denotes the mass of the fluid displaced by the particle, and ν is the kinematic viscosity of the fluid of density ρ_f; $\mathbf{g}$ is the gravitational acceleration.

The derivative

$$\frac{D\mathbf{u}}{Dt} = \frac{\partial\mathbf{u}}{\partial t} + (\mathbf{u}\cdot\nabla)\mathbf{u} \tag{3}$$

is the total hydrodynamical derivative, taken along the path of a fluid element, whereas

$$\frac{d\mathbf{u}}{dt} = \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{u} \tag{4}$$

is taken along the particle's trajectory.

The first term on the right-hand side is the acceleration of the fluid element in position $\mathbf{r}(t)$ at time t and represents the force exerted on the particle by the undisturbed fluid. The second term represents the *added-mass effect*, which accounts for the fact that when the particle moves relative to the fluid, it displaces a certain amount of fluid with it; the result is that the particle behaves as if it had additional mass. The third and fourth terms represent the Stokes drag caused by the fluid's viscosity and the buoyancy force, respectively. The integral is called the *Basset-Boussinesq history term*, and arises from the fact that the vorticity diffuses away from the particle due to viscosity [10, 11]. The terms involving the factor $a^2\nabla^2\mathbf{u}$ are the so-called *Faxén corrections* [94], and they account for the spatial variation of the flow field across the particle.

Equation (2) is valid for small particles at low particle Reynolds numbers Re_p. This Reynolds number is calculated by using the particle size as the length scale, and the relative velocity between particle and neighbouring fluid as the velocity scale: $\mathrm{Re}_p = a|\mathbf{v} - \mathbf{u}|/\nu$. This implies that for Eq. (2) to be a valid approximation, the initial velocity difference between particle and fluid must be small [1]. Another condition is that the velocity difference across the particle – more precisely, the shear Reynolds number $\mathrm{Re}_\Gamma = a^2\Gamma/\nu \ll 1$, where Γ is the typical velocity gradient in the flow – must be small [1].

If the typical length over which the velocity field changes appreciably is much larger than the particle radius a, the Faxén corrections can be neglected. Since the Basset-Boussinesq history term also describes the effect of viscosity (just like the Stokes drag), in a minimal model it can also be neglected. These approximations simplify tremendously the equations of motion. The history term would be especially problematic to analyse, since it depends on the entire past history of the particle, and it means the dynamics described by Eq. (2) has an infinite-dimensional phase-space. By neglecting it, Eq. (2) is an ordinary differential equation (and not an integro-differential equation), which can be studied with the techniques presently available to dynamical systems theory.

Using these approximations, we redefine the variables by

$$\mathbf{r} \to \mathbf{r}L, \quad \mathbf{v} \to \mathbf{v}U, \quad \mathbf{u} \to \mathbf{u}U, \quad t \to \frac{L}{U}t,$$

where L and U are the typical length and velocity scales of the flow. The new variables are all dimensionless. In these new variables, we get the following dimensionless equation of motion:

$$\ddot{\mathbf{r}}(t) = \frac{1}{\mathrm{St}}\left(\mathbf{u}(\mathbf{r}(t),t) - \dot{\mathbf{r}}(t) + W\mathbf{n}\right) + \frac{3}{2}R\frac{D}{Dt}\mathbf{u}(\mathbf{r}(t),t), \tag{5}$$

where **n** is a vertical unit vector pointing downwards, and the dimensionless parameters are

$$\mathrm{St}^{-1} = \frac{6\pi a \rho_f \nu L}{\left(m_p + \frac{1}{2} m_f\right) U}, \quad R = \frac{m_f}{m_p + \frac{1}{2} m_f}, \quad W = \frac{m_p - m_f}{6\pi a \nu \rho_f U} g, \tag{6}$$

with $g = |\mathbf{g}|$. The parameter St measures the damping intensity and is called the *Stokes number*, the dimensionless decay time in the velocity difference between particle and fluid due to the Stokes drag. The limit of $\mathrm{St} \to 0$ corresponds to the case of point particles with no inertia (since m_f and m_p are proportional to a^3). It is in this limit that the passive tracer equation (1) holds. R is the mass ratio parameter. $R < \frac{2}{3}$ corresponds to aerosols (particles heavier than the fluid), and $R > \frac{2}{3}$ corresponds to bubbles (particles lighter than the fluid). W is the scaled particle settling velocity for still fluid. Note that W/St is the dimensionless buoyancy force, which is independent of the particle size. Unless otherwise noted, we shall use Eq. (5) for the remainder of this work to describe the dynamics of finite-size particles.

2.2 General Features of the Dynamics of Finite-Size Particles

Finite-size particles have very different dynamics from that of passive tracers, which follow the same dynamics as fluid parcels. Mathematically, this is expressed by the fact that Eq. (5) is a second-order differential equation, compared with Eq. (1) which is of first order. An immediate consequence of this is that the finite-size dynamics given by Eq. (5) possesses a $2n$-dimensional phase-space, where n is the dimension of the configuration space. Thus, in a planar flow a finite-size particle is described by a dynamical system with four degrees of freedom. In contrast, for non-inertial particle dynamics the phase-space is two-dimensional.

Another difference is that the finite-size dynamics is dissipative, even in incompressible flows, and the phase-space volume contracts at the rate n/St, which is always positive. Contrast this to the non-inertial case, in which the phase-space volume coincides with the spatial volume, which renders the dynamics conservative for incompressible flows. The dissipative character of the finite-size dynamics raises the possibility of the existence of attractors in phase-space, which is not possible in the non-inertial case. This has crucial consequences for the global dynamics of particles in chaotic flows, as we shall see in later sections.

The density of an advected particle relative to the surrounding fluid plays a crucial role in its dynamics. This is incorporated in the parameter R in Eq. (5). Using a perturbative analysis valid in the limit of small particle sizes, Maxey has shown [12] that if particles have higher density than the fluid (aerosols, with $R < \frac{2}{3}$), they tend to move away from regions of high vorticity, such as the centres of eddies. This effect can be intuitively understood as the result of a centrifugal force acting on the particle and pushing it away from a highly-rotating region. Conversely, particles with lower density than the fluid (bubbles, with $R > \frac{2}{3}$) tend to move towards high-vorticity regions. So bubbles tend to agglomerate in the centres of vortices. These effects are totally absent in the non-inertial case.

3 Chaotic Advection of Passive Tracers

In order to understand properly the dynamics of finite-size particles in chaotic flows, we must first understand the simpler dynamics of passive tracers. In this section we review passive advection of particles from the viewpoint of dynamical-systems theory. In the approximation of passive advection, the particles are considered to be massless and of negligible size. They take on the velocity of the fluid flow instantaneously, and their motion is given by Eq. (1). The flows we shall consider here are typically laminar, i.e., the velocity field **u** is assumed to be smooth, although time-dependent. We briefly discuss the case of turbulent flows in the concluding Section.

3.1 Properties of Passive-Tracer Chaotic Advection

The velocity field of the fluid, described by the right-hand side of (1), is typically a non-linear function of the position and time. This implies that even if the flow itself is relatively simple and non-turbulent, the solutions of (1) can become chaotic, a phenomenon named chaotic advection by Aref [4]. This is an essential difference between the Eulerian description of fluid motion, which is concerned with the properties of the velocity field of the fluid, and the Lagrangian description, which is concerned with the trajectories of the fluid elements. It is argued in [13] that the advection in any flow with a fluid Reynolds number high enough to generate a time dependent velocity field around an obstacle displays chaos.

The stretching and folding action of the chaotic dynamics acting on a set of initial conditions in the phase space, which coincides with the configuration space, can be seen directly through the behaviour of a blob of dye injected into the fluid. As a

Fig. 1 Surface patterns downstream of the flow-through of a power plant. Loch Faskally, Scotland, photo taken by Gy. Károlyi. The scale is approximately 2 m

result, an initially compact blob of particles will trace out a complex filamentary structure; a real-world example is shown in Fig. 1. In most examples in the area of dynamical systems, these complex patterns are hidden in a high-dimensional abstract phase space; here they become visible to the naked eye, and can be photographed in experiments [14, 15].

We assume throughout this work that the fluid is incompressible. As a consequence, the motion of passively advected particles is volume preserving, hence the motion of passive tracers is similar to Hamiltonian dynamics, where the phase space volume is conserved during motion [16]. This property is unique to passively advected particles.

Advection in *two-dimensional* ($\mathbf{r} = (x,y)$, $\mathbf{u} = (u_x,u_y)$) and *incompressible* flows represents an important subclass of chaotic advection systems. Incompressibility implies that there exists a stream function $\Psi(x,y,t)$ so that the velocity components can be written as

$$u_x(x,y,t) = -\frac{\partial \Psi(x,y,t)}{\partial y}, \quad u_y(x,y,t) = \frac{\partial \Psi(x,y,t)}{\partial x}. \tag{7}$$

Substituting this into (1) we obtain the equation of motion for a particle advected in 2D in terms of the stream function:

$$\frac{dx}{dt} = -\frac{\partial \Psi(x,y,t)}{\partial y}, \quad \frac{dy}{dt} = \frac{\partial \Psi(x,y,t)}{\partial x}. \tag{8}$$

These equations have a clear Hamiltonian structure, variable x playing the role of the position, y playing the role of the conjugate momentum, and $\Psi(x,y,t)$ playing the role of the Hamiltonian function [17, 18].

If the flow is stationary, that is, the stream function Ψ does not depend explicitly on time, the particle trajectories coincide with the level curves of Ψ, called streamlines [17, 18]. From a dynamical point of view, passive advection in a stationary, incompressible 2D flow is a one-degree-of-freedom Hamiltonian system, which is always integrable. In most realistic situations, however, Ψ is not independent of time, in which case we have a one-degree-of-freedom system with a time-dependent Hamiltonian. Such driven systems typically exhibit chaotic motion [19, 20]: the advected particles move in an unpredictable way and display a great sensitivity to initial conditions. Even a very simple time dependence, for example periodicity, is enough to generate chaotic particle motion: no turbulence is necessary for complex particle trajectories. This phenomenon is often called *Lagrangian turbulence*, in contrast to the spatiotemporal complexity of a flow field, which is referred to as Eulerian turbulence.

Flows can be divided into two main classes: they can be either *open* or *closed*. A flow is closed if it is confined within a bounded domain. A flow is considered open if there is a net current flowing through the region of observation and if fluid elements cannot return there from the outflow region. A typical example of closed flow is mixing in a batch reactor without in- or outflow; an example of an open flow is the fluid motion in a channel or river in the presence of an

obstacle. In open flows, most trajectories are unbounded, and most particles escape the observation region in a finite time. In closed flows, the particles cannot escape the bounded region. Therefore chaotic behaviour, if it occurs, is persistent in closed flows. Such flows reveal structures commonly found in usual Hamiltonian systems, with chaotic regions coexisting with regular islands. Advected particles, if they start from an initial position in the chaotic region, roam the whole chaotic region, whereas particles within a regular island remain inside forever. The boundaries of these islands, formed by Kolmogorov-Arnold-Moser (KAM) tori, are impenetrable to outside particles. In the chaotic regions, particles initiated close to each other deviate exponentially along the unstable foliation of unstable fixed points. This unstable foliation is dense in the chaotic regions, and governs the stretching and folding of dye blobs. When stretching of a dye blob starts, a filamentary structure emerges, which becomes space-filling in the limit of long times. Thus, *persistent chaos* and *transient filamentary structures* are features found in closed flows.

The chaotic motion of passively advected particles in open flows takes a different form. Typical particles escape the observation region in finite time, but there is a fractal set of particle trajectories confined within a finite region, e.g. in the vicinity of the wake in a flow around an obstacle. These non-escaping orbits form a non-attracting chaotic set, a *chaotic saddle* [19, 20], which governs the motion of particles in its vicinity. This chaotic set, although it consists of unstable orbits and has measure zero, gives rise to extreme sensitivity of the dynamics to initial conditions. The stable manifold of the chaotic saddle separates the initial conditions leading to different final states of the particles (such as where they leave the region of observation). The unstable manifold of the chaotic saddle is traced out by the particles that spend long times in the vicinity of the saddle. Because the unstable manifold is a filamentary fractal, the pattern traced out by a blob of dye takes on a complex shape, shadowing the fractal structure of the unstable manifold. Advection thus leads to *transient chaos* owing to particles leaving the region of observation, and to *persistent filamentary structures* traced out by the advected particles.

In the following subsections some paradigmatic flows, both open and closed, are reviewed. These flows will be used later to illustrate the motion of finite-size particles.

3.2 The Convection and Cellular Flow Models

The convection flow is a simple two-dimensional incompressible flow representing vortices or roll cells with oscillating velocity magnitude. It was introduced by Chandrasekhar [21] as a solution to the Bénard problem and has been used since then in many studies involving active and passive particles [22, 23]. The flow is defined on a unit cell, but one may study the motion of finite-size particles on an infinite spatial domain using periodic boundary conditions. The convection flow is particularly interesting to investigate the principles of particle motion in the ocean and in the atmosphere since it contains both vortices (convection cells) and linear

uprising/sinking regions, and can hence be considered as an idealisation of realistic atmospheric and oceanic flows. The flow is given by the stream function

$$\psi(x,y,t) = [1 + B\sin(\omega t)]\frac{1}{k}\sin(kx)\sin(ky). \tag{9}$$

Here it is assumed that the characteristic velocity $U_0 = 1$ and the characteristic length scale of the flow $L = 1$, and we set $k = 2\pi$ accordingly. The parameters B and ω denote the amplitude and the frequency of the oscillation of the vortical velocity magnitude, respectively.

A special feature of the flow described by Eq. (9) it that the non-inertial advection dynamics remains nonchaotic even in the time-dependent case. In this flow chaotic advection can only be the consequence of finite-size effects.

To illustrate the flow field we show a snapshot of the velocity field in Fig. 2. This flow field with the parameters $B = 2.72$, $\omega = \pi$ is used later for all results concerning coagulation and fragmentation processes (cf. Sect. 6).

We also consider a slightly modified version of the convection flow, the cellular flow model. This is a two-dimensional incompressible flow representing a lattice of oscillating vortices or roll cells. The flow is defined by the stream function

$$\psi(x,y,t) = \frac{1}{k}\cos(kx + B\sin(\omega t))\cos(ky). \tag{10}$$

The flow is defined in the $x,y \in [-\pi/2,\pi/2]$ domain with periodic boundary conditions, and k is set to 1.

Let us first consider the simplest case where the time dependence is suppressed, by setting $B = 0$. Thence ψ is a constant of motion, which implies that real fluid elements follow trajectories that are level curves of ψ. This is illustrated in Fig. 3a. For $B \neq 0$, the equations of motion for fluid element or passive tracer are given by (7) and the trajectories differ from the streamlines. An example is shown in Fig. 3b, where the Poincaré section of the particle motion is illustrated. The positions of many particles are plotted at integer multiples of the flow period. Some of the

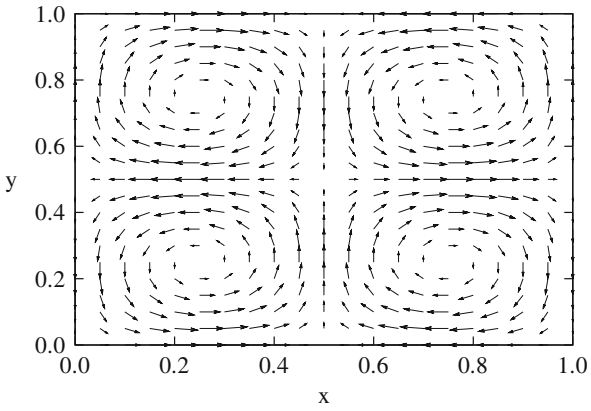


Fig. 2 Snapshot of the velocity field of the convection flow at $t = 0$ computed from the stream function (9)

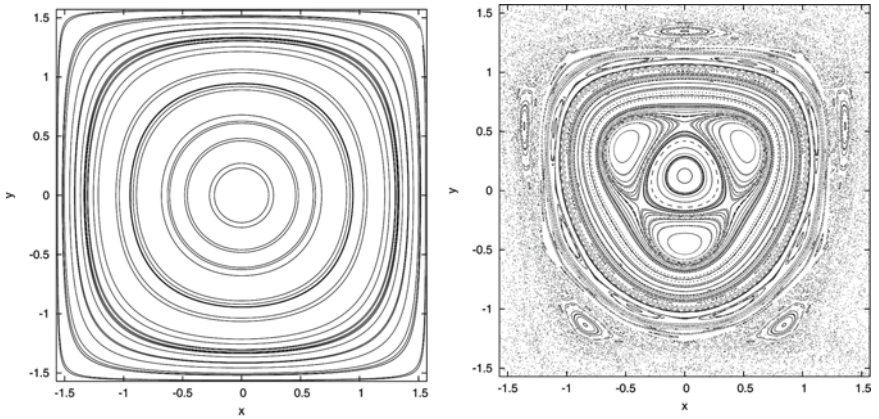


Fig. 3 The cellular flow, Eq.(10). (**a**) The passively advected particles exhibit regular motion if the flow is time-independent ($B = 0$, $k = 1$). The trajectories of 30 particles with randomly selected initial positions are shown. (**b**) Chaotic behaviour in the time-dependent flow ($B = 0.3$, $k = 1$, $\omega = 2.72224$). Snapshots of 200 particles with randomly selected initial conditions are shown at integer multiples of the flow's period

passively advected particles exhibit regular motion; they trace out the closed curves visible in the figure. The rest of the particles fill out the chaotic region. This makes the cellular flow different from the convective flow in that the passive particle motion in the cellular flow, due to the oscillating vortex centres, can be chaotic for $B \neq 0$.

3.3 The Von Kármán Vortex Street

The open flow around an obstacle is a classical problem in fluid mechanics [14, 24]. We consider a viscous incompressible flow around a cylinder of radius R_0. Far away from the obstacle the flow is expected to be uniform. We label the longitudinal flow direction by x, and the transverse direction by y.

Denoting by U the velocity for $x \to \pm\infty$, the Reynolds number associated with this flow can be defined as

$$\mathrm{Re} = 2R_0U/\nu, \tag{11}$$

where ν is the fluid's kinematic viscosity. For Re sufficiently small, the flow is stationary. When Re passes a critical value $\mathrm{Re}_c \approx 80$, the stationary solution of the Navier-Stokes equation becomes unstable, and the flow becomes time-periodic with some period T.

Vortices are created in the wake of the cylinder, detach from it and drift downstream. They gradually weaken owing to the viscosity, until after some distance they vanish. New vortices are shed from the surface of the cylinder at intervals of half a

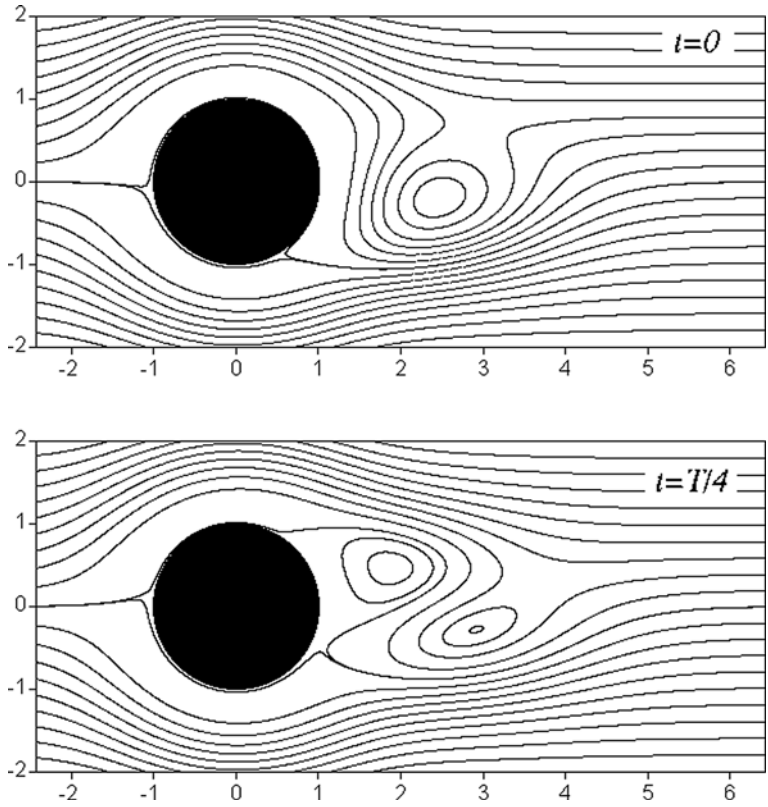


Fig. 4 *Streamlines* for the flow around a cylinder at two different times, separated by one quarter of the full period T of the flow. The vortex shedding is clearly visible

period $T/2$, alternately above and below the middle of the cylinder (see Fig. 4). By this process, a *von Kármán vortex street* is formed behind the cylinder. For simplicity we assume that the lifetime of each detached vortex equals one period T.

An analytical model for the flow in the von Kármán vortex street has been proposed [25, 26], which fits well the results of the direct numerical calculation for $\mathrm{Re} = 250$ [27]. In this model, the stream function $\psi(x,y,t)$ is explicitly given, and we shall use this kinematic model in what follows. This model serves thus as an ideal paradigm for a large class of open chaotic flows and has been widely used to study different aspects of transient chaotic advection (see e.g., [28–30]).

The stream function $\psi(x,y,t)$ can be directly used in Eq. (1) to find the motion of passively advected particles in the von Kármán flow by numerical integration. To appreciate the importance of the unstable manifold of the chaotic saddle, we place a dye droplet of particles upstream into the flow and follow the deformation of the shape of this droplet. Assume that the initial droplet overlaps with the stable manifold of the saddle. Particles that fall exactly on the stable manifold hit the saddle and never leave it. Neighbouring points, however, only approach the saddle; they stay

in its neighbourhood for a while, but sooner or later they leave the wake along the unstable manifold. Thus, we conclude that tracer particles that do not leave a region of observation in the wake too rapidly, must trace out the unstable manifold. In other words, the unstable manifold of the chaotic saddle is a "quasi-attractor" of the tracer dynamics: particles accumulate on it while being advected away. In numerical simulations with a finite number of particles, the manifold serves as a (periodically moving) template, which becomes gradually emptied as more and more particles escape through the outflow.

Figure 5 shows the evolution of a droplet in the von Kármán flow. First the droplet becomes stretched and folded and later it becomes clear that it traces out a moving fractal object, the unstable manifold. Though a few particles are still visible, the region is almost emptied in the last panel; this is a consequence of the finite number of particles used in the simulation.

Note that the von Kármán vortex street is not a particular property of cylindrical obstacles. Most (approximately) two-dimensional flows past an obstacle have this property, provided that their Reynolds number is in the appropriate range. Thus, von Kármán vortices are found in many real situations [31–33].

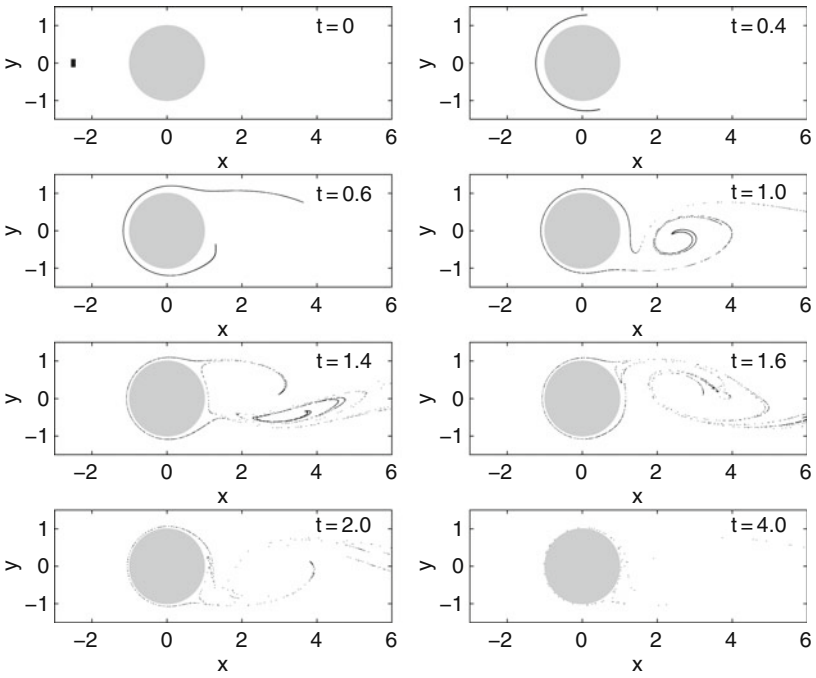


Fig. 5 Time evolution of a droplet of 20,000 tracers in the von Kármán flow shown at different dimensionless times t

4 Inertial Effects in Closed Chaotic Flows

The most general description of the dynamics of finite-size particles presents an enormous richness of phenomena (see, for example, [7, 9, 34–40]). It is characteristic of the dynamics that invariant surfaces in the model without inertia are broken up. For example, as shown for aerosols in Fig. 6, the particles accumulate on higher dimensional attractors instead of being confined by the closed curves shown in Fig. 3b. Figure 6 shows the 2D projection of the attractor located in the four-dimensional phase space of the particle dynamics.

That the invariant curves no longer exist for inertial, finite-size particles is confirmed in Fig. 7. Here, the periodic boundary conditions have been removed from the cellular flow, and the particles are allowed to fly out of the $x,y \in [-\pi/2,\pi/2]$ domain. For passive advection (not shown), the $y = \pm\pi/2$ lines are impenetrable invariant curves; the advected particles can only leave the $x,y \in [-\pi/2,\pi/2]$ domain in the x direction. As shown in Fig. 7 the invariant curves $y = \pm\pi/2$ do not exist for aerosols, which can fly out in the y direction as well.

In the time-independent case, what were invariant surfaces in the model without inertia are transformed into spirals, owing to centrifugal forces: outward spirals for aerosols and inward spirals for bubbles. As a consequence, heavy particles tend to accumulate at the separatrices of the flow.

For large densities $\rho_p \gg \rho_f$, particles are no longer confined within vortices. Stokes drag is the most important force acting in this case, since the added-mass term becomes negligible as $R \to 0$ in this limit (see Eq. (6)) and gravity is not acting on the horizontal plane. So to a first approximation Eq. (5) transforms into

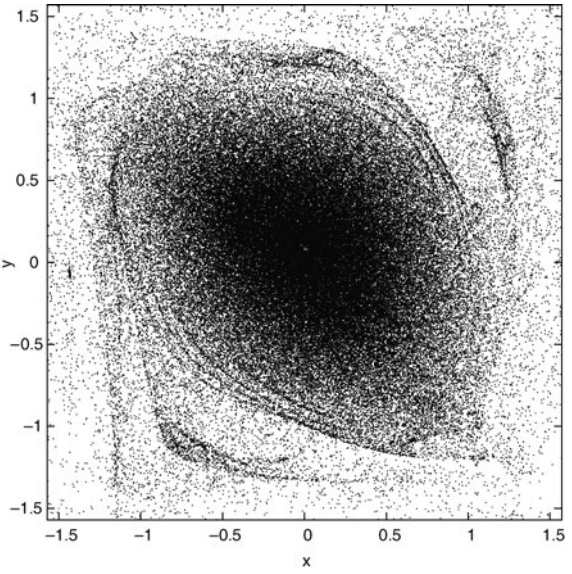


Fig. 6 Aerosols ($R = 0.5$, $\mathrm{St}^{-1} = 0.04$) followed in the cellular flow. The particle positions are shown at integer multiples of the flow's period after 10 periods. The parameters of the flow are the same as those in Fig. 3b

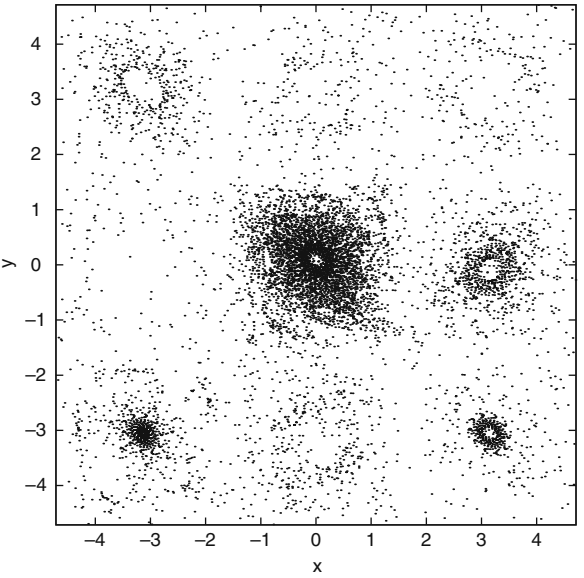


Fig. 7 Aerosols ($R = 0.5$, $\mathrm{St}^{-1} = 0.04$) followed in the cellular flow without periodic boundary conditions. All particles are initiated within the $x,y \in [-\pi/2, \pi/2]$ domain. The particle positions are shown at integer multiples of the flow's period. The parameters of the flow are the same as those in Fig. 3b

$$\frac{d^2x}{dt^2} = -\mathrm{St}^{-1}\left(\frac{dx}{dt} - u_x(x,y,t)\right), \tag{12}$$

$$\frac{d^2y}{dt^2} = -\mathrm{St}^{-1}\left(\frac{dy}{dt} - u_y(x,y,t)\right), \tag{13}$$

and the Stokes number can be written as

$$\mathrm{St}^{-1} = \frac{9\nu\rho_f L}{2a^2\rho_p U}.$$

To derive this we used $m_p = \frac{4}{3}\pi a^3 \rho_p \gg m_f$ in Eq. (6).

This is a highly-dissipative and singular perturbation of a Hamiltonian system, with a four-dimensional phase space:

$$\dot{x} = p_x, \tag{14}$$

$$\dot{p}_x = -\mathrm{St}^{-1}(p_x - u_x(x,y,t)), \tag{15}$$

$$\dot{y} = p_y, \tag{16}$$

$$\dot{p}_y = -\mathrm{St}^{-1}(p_y - u_y(x,y,t)). \tag{17}$$

In the time-dependent case, particles tend to accumulate on a chaotic attractor of the high-dimensional phase space. As the upper panel of Fig. 8 shows for the cellular flow (9), the projection of the attractor on the plane of the fluid lies within the chaotic regions of the model flow without inertia. The relative velocity fluctuates

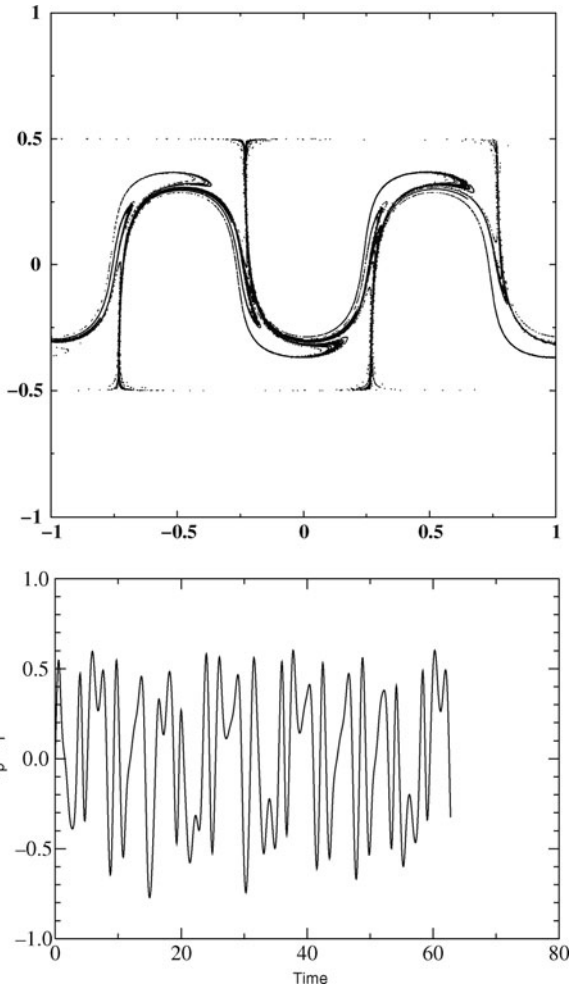


Fig. 8 *Above*: dense particles converge to fractal structures around a separatrix of the flow given by Eq. (10) in the time-independent case. *Below*: in the large inertia limit, the relative velocity of particle and flow fluctuates chaotically, in a Brownian-like fashion

chaotically, due to macroscopic, non-turbulent fluctuations, that act to give the particles deterministic but Brownian-like motion, illustrated in the lower panel of Fig. 8.

4.1 Neutrally Buoyant Particles

Let us now consider whether even in the most favourable case of neutral buoyancy a finite-sized tracer particle remains always close to a flow trajectory [41]. With this in mind, we set $\rho_p = \rho_f$ in Eqs. (5) and (6), which corresponds to setting $W = 0$ and $R = 2/3$:

$$\frac{d\mathbf{v}}{dt} = \frac{D\mathbf{u}}{Dt} - \mathrm{St}^{-1}\,(\mathbf{v} - \mathbf{u})\,. \tag{18}$$

In the past it has been assumed that neutrally buoyant particles have trivial dynamics (e.g., [34, 42]), and the mathematical argument used to back this up is that if we make the approximation $D\mathbf{u}/Dt = d\mathbf{u}/dt$, which can be seen as a rescaling of the added mass, the problem becomes very simple

$$\frac{d}{dt}(\mathbf{v} - \mathbf{u}) = -\mathrm{St}^{-1}(\mathbf{v} - \mathbf{u}). \tag{19}$$

Thence

$$\mathbf{v} - \mathbf{u} = (\mathbf{v}_0 - \mathbf{u}_0) \exp(-\mathrm{St}^{-1}\, t), \tag{20}$$

from which we would infer that even if we release the particle with a different initial velocity $\mathbf{v}_0$ to that of the fluid $\mathbf{u}_0$, after a transient phase the particle velocity will match the fluid velocity, $\mathbf{v} = \mathbf{u}$, meaning that if we accept this argument, a neutrally buoyant particle should be an ideal tracer.

Although from the foregoing it would seem that neutrally buoyant particles represent a trivial limit to Eq. (2), in the argument presented above we did not take the correct approach to the problem, because we did not recognise that $D\mathbf{u}/Dt \neq d\mathbf{u}/dt$. If we substitute Eqs. (3) and (4) for the derivatives into Eq. (18), we obtain

$$\frac{d}{dt}(\mathbf{v} - \mathbf{u}) = -((\mathbf{v} - \mathbf{u}) \cdot \nabla)\,\mathbf{u} - \mathrm{St}^{-1}(\mathbf{v} - \mathbf{u}). \tag{21}$$

We may then write the velocity difference between fluid particle and fluid as $\mathbf{q} = \mathbf{v} - \mathbf{u}$, whence

$$\frac{d\mathbf{q}}{dt} = -\left(-J + \mathrm{St}^{-1} I\right) \cdot \mathbf{q}, \tag{22}$$

where J is the Jacobian matrix:

$$J = \begin{pmatrix} \partial_x u_x & \partial_y u_x \\ \partial_x u_y & \partial_y u_y \end{pmatrix}. \tag{23}$$

If we diagonalise matrix J we obtain the equation for the particle-fluid velocity difference in coordinates aligned with the eigenvectors, which we denote by $\mathbf{q}_D$:

$$\frac{d\mathbf{q}_D}{dt} = \begin{pmatrix} \lambda - \mathrm{St}^{-1} & 0 \\ 0 & -\lambda - \mathrm{St}^{-1} \end{pmatrix} \cdot \mathbf{q}_D. \tag{24}$$

Therefore, if $\mathrm{Re}(\lambda) > \mathrm{St}^{-1}$, $\mathbf{q}_D$ may grow exponentially. Now λ satisfies $\det(J - \lambda I) = 0$, so $\lambda^2 - \mathrm{tr}J + \det J = 0$. Since the flow is incompressible, $\partial_x u_x + \partial_y u_y = \mathrm{tr}J = 0$, thence $-\lambda^2 = \det J$. Given squared vorticity $\omega^2 = (\partial_x u_y - \partial_y u_x)^2$, and squared strain $s^2 = s_1^2 + s_2^2$, where the normal component is $s_1 = \partial_x u_x - \partial_y u_y$ and the shear component is $s_2 = \partial_y u_x + \partial_x u_y$, we may write

$$\lambda^2 = -\det J = (s^2 - \omega^2)/4 = Q. \tag{25}$$

Here $Q = (s^2 - \omega^2)/4$ is the so-called Okubo–Weiss parameter [43, 44]. If $Q > 0$, $\lambda^2 > 0$, and λ is real, deformation dominates, as around hyperbolic points. If $Q < 0$, $\lambda^2 < 0$, and λ is complex, rotation dominates, as near elliptic points. Equation (22) together with $d\mathbf{r}/dt = \mathbf{q} + \mathbf{u}$ defines a dissipative dynamical system

$$d\boldsymbol{\xi}/dt = \mathbf{F}(\boldsymbol{\xi}) \tag{26}$$

with $\boldsymbol{\xi} = (\mathbf{r},\mathbf{q})$. Equation (26) has constant divergence $\nabla \cdot \mathbf{F} = -2/\mathrm{St}$ in the four dimensional phase space of $\boldsymbol{\xi}$. While small values of St allow for large values of the divergence, large values of St force the divergence to be small. The Stokes number is the dimensionless decay time of the particle (see Sect. 2.1): with larger St, the particle has more independence from the fluid flow. From Eq. (24), in areas of the flow near hyperbolic stagnation points with $Q > \mathrm{St}^{-2}$, particle and flow trajectories separate exponentially.

To illustrate the effects of St and Q on the dynamics of a neutrally buoyant particle, let us consider the simple incompressible two-dimensional model (10). In Fig. 9 (top left) the contours of Q are depicted. Notice that the high values of Q are around the hyperbolic points, while negative Q coincides with the centres of vortices – elliptic points – in the flow. Figure 9 (top right) shows the trajectory of a neutrally buoyant particle starting from a point on a fluid trajectory within the central vortex, but with a small velocity mismatch with the flow. This mismatch is amplified in the vicinity of the hyperbolic stagnation points where Q is larger than St^{-2} to the extent that the particle leaves the central vortex for one of its neighbours. In the end a particle settles on a trajectory that does not visit regions of high Q, as expected for a fluid parcel. While this effect is already seen in Fig. 9 (top right), it is more dramatically pictured in the trajectory shown in Fig. 9 (bottom left), in which the particle performs a long and complicated excursion wandering between different vortices before it settles in a region of low Q. To illustrate the divergence of particle and fluid trajectories, and the fact that particle and fluid finally arrive at an agreement, in Fig. 9 (bottom right) we display the difference between the particle velocity and the fluid velocity at the site of the particle against time for this case. Notice that this difference seems negligible at time zero, and that it also convergences to zero at long times, but during the interval in which the excursion takes place it fluctuates wildly.

Even more interesting is the case of time-dependent flows: $B \neq 0$ in our model. As in a typical Hamiltonian system, associated with the original hyperbolic stagnation points, there are regions dominated by chaotic trajectories. Trajectories of this kind, stroboscopically sampled at the frequency of the flow, are reproduced in Fig. 10. Such trajectories visit a large region of the space, which includes the original hyperbolic stagnation points and their vicinities where Q is large. Excluded from the reach of such a chaotic trajectory remain areas where the dynamics is regular: KAM tori. In our model these lie in the regions where $Q < \mathrm{St}^{-2}$. A neutrally buoyant particle trying to follow a chaotic flow pathline would eventually reach the highly

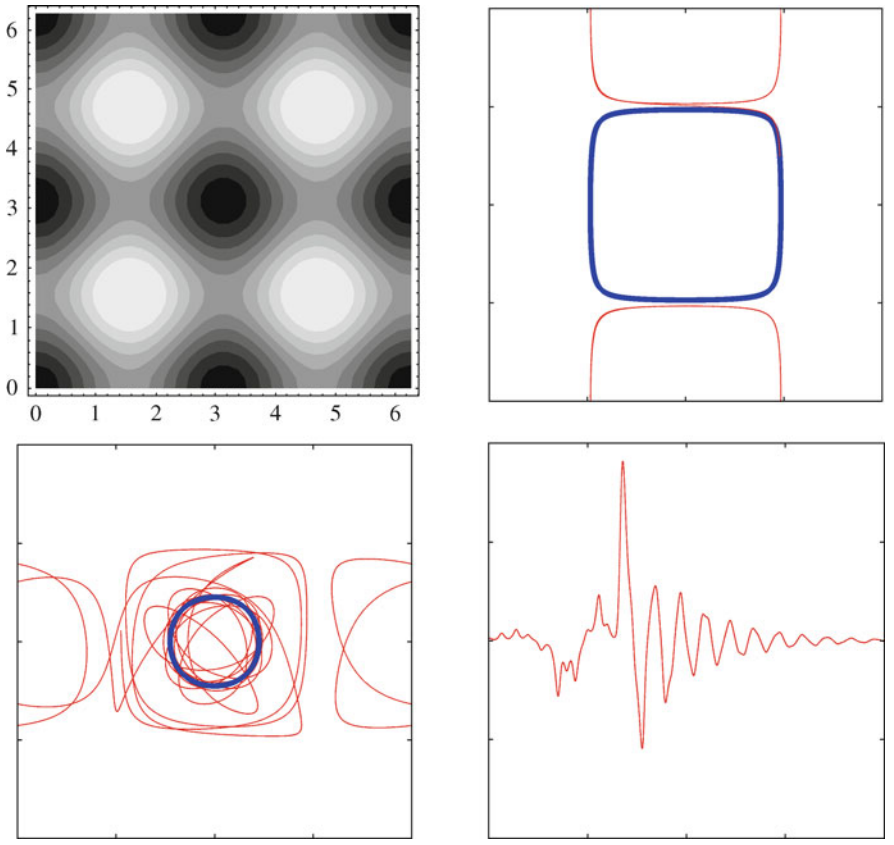


Fig. 9 (*Top left*) Contour plot illustrating magnitude of Q – lighter is higher Q – for the time-independent model Eq. (10) (the flow is on a torus). (*Top right*) The separation of a neutrally buoyant particle trajectory (*thin line*) from the flow (*thick line*) in regions of high Q allows the particle to wander between cells. (*Bottom left*) After a complicated excursion, a particle (*thin line*) eventually settles in a zone of low Q of the flow; a KAM torus (*thick line*). (*Bottom right*) The velocity difference $v_x - u_x$ between the particle and the flow against time

hyperbolic regions of the flow. This makes likely its separation and departure from such a pathline, in search of another pathline to which to converge. However, convergence will only be achieved if the pathline never crosses areas of high Q. Figure 10 demonstrates this phenomenon: a particle was released in the chaotic zone with a small velocity mismatch. The particle followed the flow, until, coming upon a region of sufficiently high Q, it was thrown out of that flow pathline onto a long excursion that finally ended up in a regular region of the flow on a KAM torus. The regular regions of the flow then constitute attractors of the dissipative dynamical system Eq. (26) that describes the behaviour of a neutrally buoyant particle. The chaotic trajectories in a Hamiltonian system are characterised by positive Lyapunov exponents. The Lyapunov exponents are an average along the trajectory of the local rate

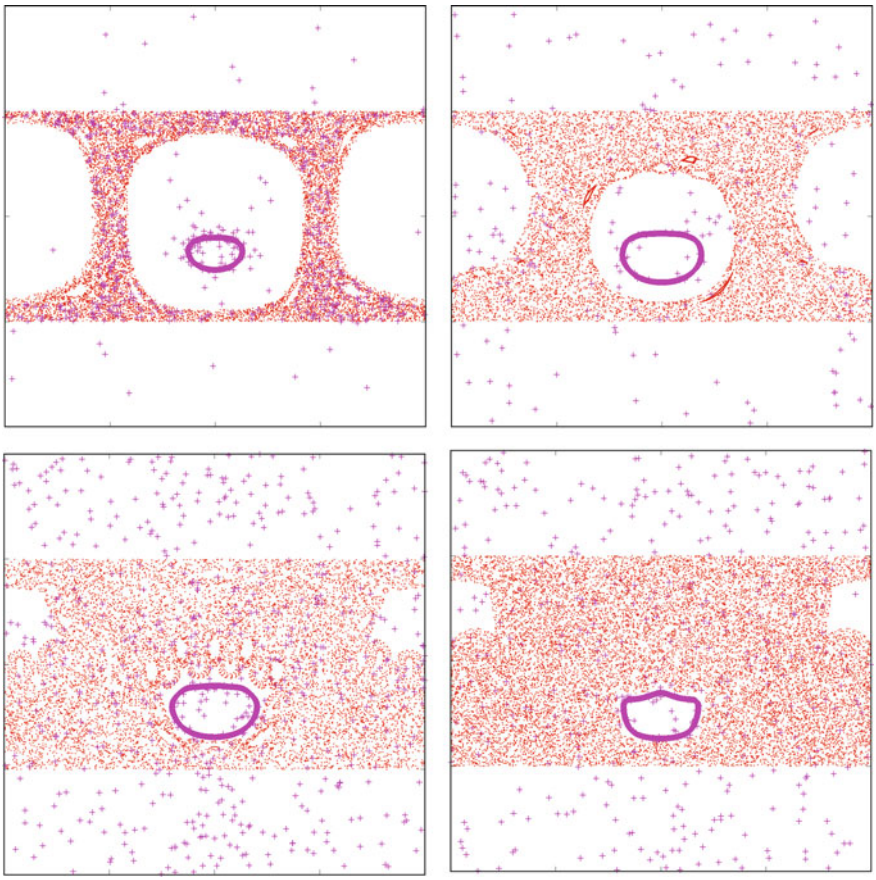


Fig. 10 Poincaré sections of trajectories in the time-dependent flow of Eq. (10). From *top left* to *bottom right* are shown (*dots*) four increasingly chaotic examples of the flow, and (crosses), the trajectories of neutrally buoyant particles in the flows that in each case finally end up on a KAM torus within the regular region of the flow

of convergence or divergence. Hence, for a trajectory to be chaotic, it is a necessary condition that it visit regions of positive Q: an upper bound to Q is an upper bound to the Lyapunov exponent.

Consider the implications of these results for two-dimensional turbulent flows, in which Q defines three regions: in the vortex centres it is strongly negative; in the circulation cells that surround them, strongly positive, while in the background between vortices it fluctuates close to zero (see, e.g., [45–49]). As a result of the dynamics, an initially uniform distribution of neutrally buoyant particles with finite size evolves in time towards an asymptotic distribution concentrated in the inner part of vortices where $Q < 0$, and with voids in the areas crossed by fluid trajectories that visit regions where $Q > \mathrm{St}^{-2}$, as we illustrate in Fig. 11.

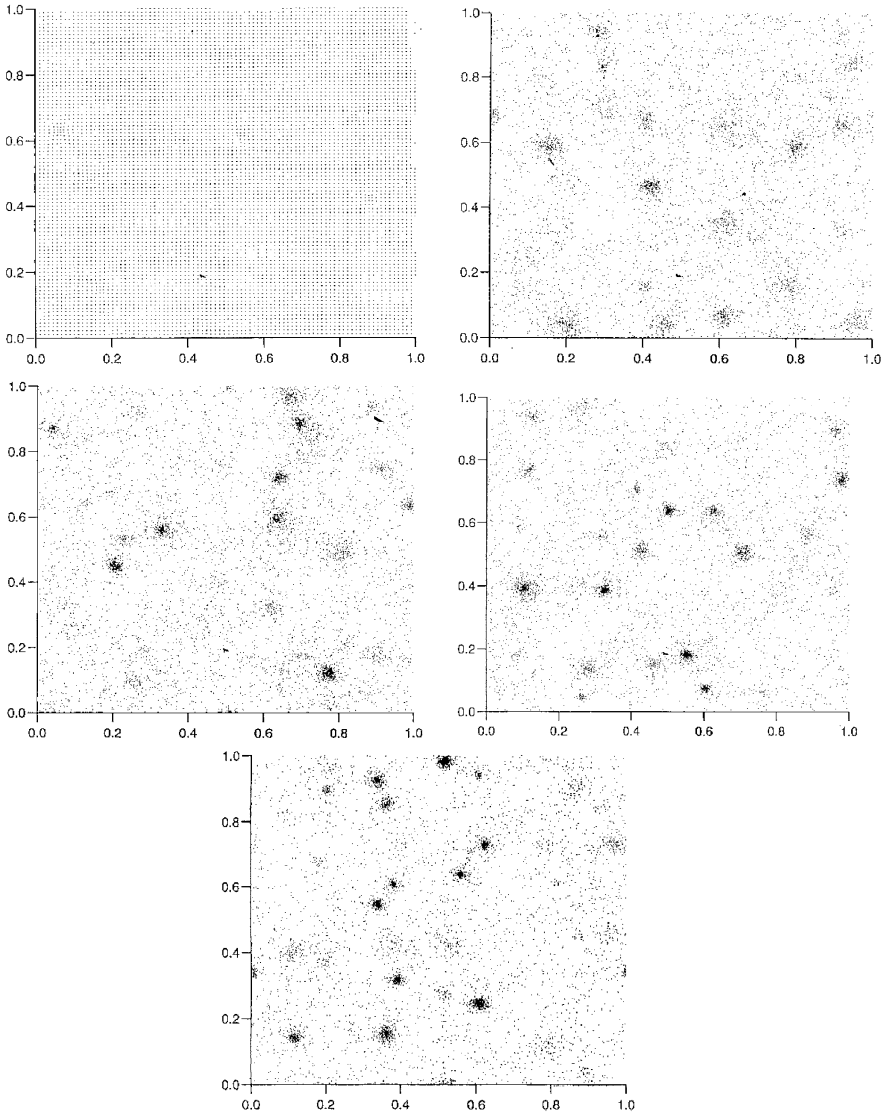


Fig. 11 Small neutrally buoyant tracer particles converge to the centres of vortices in a two-dimensional turbulent flow simulation; distribution at times (*top left* to *bottom right*) $t = 1$, 2, 3, 4, and 6 of particles uniformly distributed in the flow at time $t = 1$ [41]

Thus even with a small rigid neutrally buoyant spherical tracer particle in an incompressible two-dimensional fluid flow the tracer trajectories can separate from the fluid trajectories in those regions where the flow has hyperbolic stagnation points. For flows with chaotic pathlines, analysis shows that the tracer will only evolve on trajectories having Lyapunov exponents bounded by the value

of the Stokes drag coefficient. Therefore, by making the value of this coefficient small enough, one can force the tracer to settle on either the regular KAM-tori dominated regions or to selectively visit the chaotic regions with small Lyapunov exponents. As well as its interest from the viewpoint of dynamical systems, this result is important to the analysis of observations and experiments with neutrally buoyant particles both in the laboratory and in the atmosphere and oceans.

The heuristic criterion for the departure of trajectories of neutrally buoyant particles trajectories from those of the fluid elements has been complemented by more rigorous analysis. For example, the stability of the fluid flow manifold which is invariant under the neutrally buoyant particle dynamics was studied in the more general context of the perturbed invariant manifold of the non-neutral particles [50, 51]. For neutrally buoyant particles, this analysis parallels a previous interpretation of the fluid invariant manifold as a "bailout embedding" and the departure of trajectories as a "blowout bifurcation" taking place as the Stokes number is varied [52]. In both cases, it has been found that the heuristic criterion based on the Okubo-Weiss parameter underestimates the areas of departure. Finally noise or fluctuating forces enhance the phenomenon, as has been rigorously proved in [53].

4.2 3D Flows and Bailout Embeddings

For incompressible two-dimensional flows, since the Jacobian matrix is traceless, the two eigenvalues must add up to zero, which implies that they are either both purely imaginary or both purely real, equal in absolute value and opposite in sign. The result is that the particles can abandon the fluid trajectories in the neighbourhood of the saddle points and other unstable orbits, where the Jacobian eigenvalues are real, and eventually overcome the Stokes drag, to finally end up in a regular region of the flow on a KAM torus dominated by the imaginary eigenvalues. From a more physical point of view, this effect implies that the particles tend to stay away from the regions of strongest strain.

In contrast to the two-dimensional case, in time-dependent 3D flows the incompressibility condition only implies that the sum of the three independent eigenvalues must be zero. This less restrictive condition allows for many more combinations. Triplets of real eigenvalues, two positive and one negative or vice versa, as well as one real eigenvalue of either sign together with a complex-conjugate pair whose real part is of the opposite sign, are possible. Accordingly, chaotic trajectories may have one or two positive Lyapunov numbers, and a richer range of dynamical situations may be expected.

Note that the dynamical system governing the behaviour of neutrally buoyant particles is composed of a lower-dimensional dynamics within a "larger", higher-dimensional dynamics. Equation (18) can be seen as an equation for the variable $\mathbf{q} = (\mathbf{v} - \mathbf{u})$ which in turn defines the equation of motion ($\dot{\mathbf{r}} = \mathbf{u}$) of a fluid

element whenever the solution of the former is zero. In this sense we may say that the fluid parcel dynamics is embedded in the particle dynamics. In reference to the fact that some of the embedding trajectories abandon some of those of the embedded dynamics, the generalisation of this process is dubbed a *bailout embedding* [52].

Consider neutrally buoyant particles immersed in a flow in which each component of the velocity vector field is sinusoidally modulated with a relative phase shift of $2\pi/3$ and where x, y, and z are to be considered $(\mathrm{mod}2\pi)$

$$\begin{aligned}
\frac{dx}{dt} &= (1 + \sin\ 2\pi t) \cdot (A \sin z + C \cos y), \\
\frac{dy}{dt} &= \left(1 + \sin\ 2\pi \left(t + \frac{1}{3}\right)\right) \cdot (B \sin x + A \cos z), \\
\frac{dz}{dt} &= \left(1 + \sin\ 2\pi \left(t + \frac{2}{3}\right)\right) \cdot (C \sin y + B \cos x).
\end{aligned} \tag{27}$$

This is a modified version of the ABC flow [54]. This flow shows structures consisting of a complex array of KAM sheets and tubes surrounded by chaotic volumes [55]. Neutrally buoyant particles show a tendency to accumulate inside KAM tubes as depicted in Fig. 12, where ten particles, initially distributed at random in the cubic cell, are shown to end up in the interior of two of the tubes.

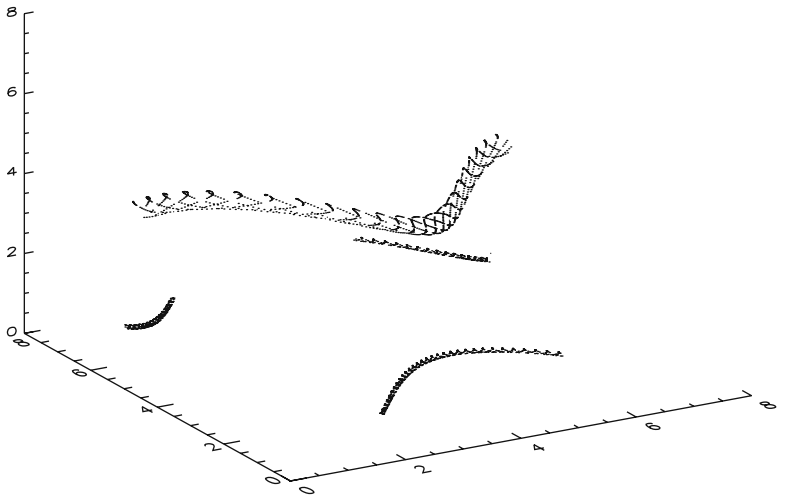


Fig. 12 Stroboscopic sampling (with period T=1) of the position of 10 particles initially distributed at random in a flow described by Eq. (27) with $A = 2$, $B = 0.4$, $C = 1.2$. The *dots* represent the positions of these particles at the strobing periods 1,000–2,000

5 Advection of Finite-Size Particles in Open Flows

Here we concentrate on the question of how the global dynamics of advection changes in two-dimensional open flows due to inertia. As discussed in Sect. 2, one big difference between finite-size and non-inertial particles is that the dynamics of the former is dissipative, which opens the possibility of the existence of attractors. In the flow model of the von Kármán vortex street of Sect. 3.3, it has been found that attractors are possible in the bubble regime $2/3 < R < 2$. Light particles might thus *become trapped* in the wake forever. For $R = 1.33$, $\mathrm{St}^{-1} = 30$ there are, for instance, three coexisting attractors [56–58]: two fixed points around the cylinder's surface and one at $x = \infty$.

To gain insight into what happens to ensembles of bubbles, the residence time in a region around the cylinder is determined. The initial velocities were set to be equal to the flow velocity. Figure 13 shows the result, where grey, white and black depict increasingly long residence times. Dark regions mark permanently trapped particles. This region corresponds thus to the basin of attraction of the two finite attractors, which is in fact a projection of the basin structure in the full phase space on the plane of the flow. At other inertia parameters there also exist chaotic saddles which ensure that the approach toward the attractors, including the escape from the wake (the approach toward the attractor at $x = \infty$) is a transient chaotic process.

A systematic investigation of the escape rate $\kappa(\mathrm{St})$ from the hyperbolic parts of this saddle shows (Fig. 14) that the escape rate is *below* the escape rate of fluid parcels or passive tracers in the full range $\mathrm{St}^{-1} > 12$. This indicates that bubbles spend much more time in the wake than fluid particles. In the interval $14 < \mathrm{St}^{-1} < 45$ the escape rate vanishes indicating the presence of attractors. For St^{-1} in between 33 and 45 these attractors are chaotic. Beyond 45 the escape rate is positive, and it approaches for large St^{-1} the value of ideal tracers.

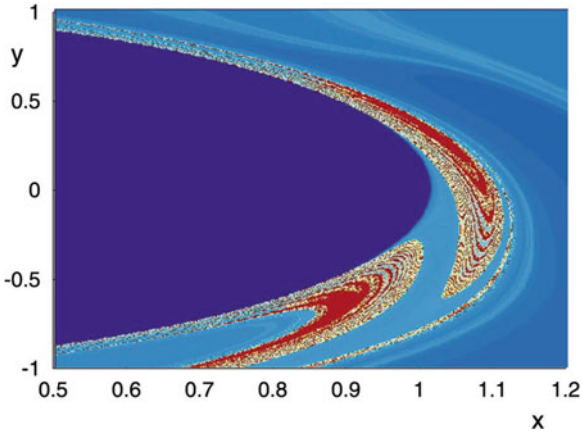


Fig. 13 Finite-size particles in the von Kármán flow. Residence time for bubbles of parameter $\mathrm{St}^{-1} = 30$, $R = 1.33$ ($g = 0$) at $t = 0.3$ mod 1. Basins of attraction of two chaotic attractors appear *shaded dark* in the plane of the fluid. From [57]

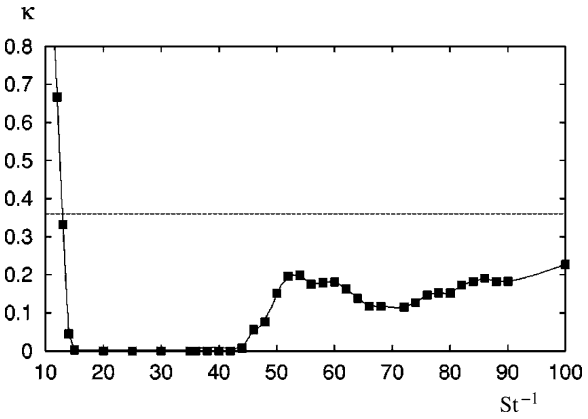


Fig. 14 Escape rate as a function of the Stokes number in the bubble regime ($R = 1.7$). The *horizontal line* is the escape rate for passive tracers. From [57]

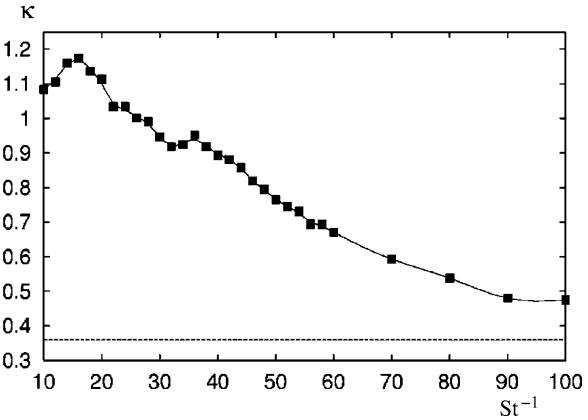


Fig. 15 Escape rate as a function of the Stokes number in the aerosol regime ($R = 0.5$). The *horizontal line* is the escape rate for passive tracers. From [57]

The tendency is opposite for aerosol particles. The escape rate is *above* the escape rate of fluid particles for any value of St, i.e. heavy particles spend much less time in the wake than fluid particles (Fig. 15).

There is a qualitative argument explaining why bubbles tend to form attractors. The particles are typically subject to local vortices. The centrifugal force for a particle comoving with a vortex is proportional to the density difference $\rho_p - \rho_f$. For heavy particles this force pushes particles outward, but for light ones it attracts particles toward the vortex centre. The presence of this centripetal force is an important reason for the existence of bubble attractors. The mechanism is similar to an observation of Maxey [12] according to which aerosols (bubbles) settle in the presence of gravity faster (slower) in turbulent flows than in a fluid at rest. The explanation of this phenomenon is the centrifugal (centripetal) effect of the turbulent vortices.

As a consequence of the dependence of the dynamics on particle parameters, when starting from a mixture of particles of the same density but of different size, segregation can be observed in chaotic flows consisting of a sequence of obstacles. As an example of this type of flow, we consider an infinite chain of cylinders situated

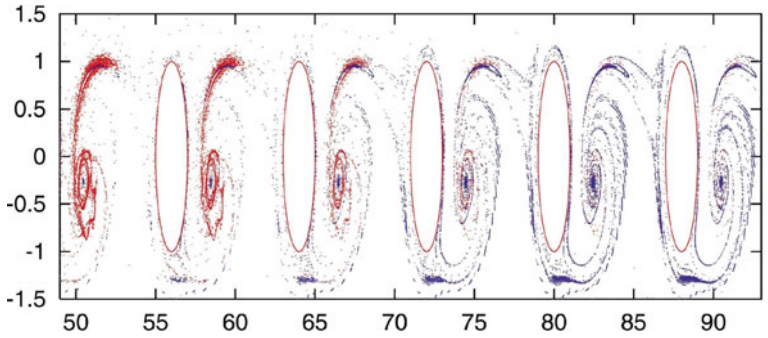


Fig. 16 Separation of bubbles in a chain of cylinders. The distribution of particles with $R = 1.4$ and $\mathrm{St}^{-1} = 20$ $\mathrm{St}^{-1} = 120$ coloured *red* and *blue*, respectively, after 10 time units. The initial location is a *small square* with uniformly mixed particles in front of the cylinder located at the origin. From [57]

at a distance of 8 cylinder radii from each other. The initial droplet of particles contains a uniform mixture of particles of $\mathrm{St}^{-1} = 20$ (red) and $\mathrm{St}^{-1} = 120$ (blue), and is injected into the flow in front of the cylinder centred at the origin. After passing several cylinders, the droplet exhibits a clear separation as shown in Fig. 16. The larger light particles ($\mathrm{St}^{-1} = 20$) escape more slowly due to the centripetal effect of the vortices. The flow in the cylinder chain acts therefore as a chaotic chromatograph [57].

A work worthy of note is that of Haller and coworkers [50, 59]. They show that for small particle sizes, i.e., for small Stokes numbers ($\mathrm{St} \ll 1$), the dynamics of a finite-size particle can be approximated by the dynamics on a low-dimensional inertial manifold, which can be calculated explicitly from a given velocity field. Following an approach of Maxey [12], they show that after a short transient time the equation of motion of such small inertial particles can well be approximated by the equation, termed inertial equation in [50]

$$\dot{\mathbf{r}}(t) = \mathbf{u}(\mathbf{r}(t),t) + W\mathbf{n} + \mathrm{St}\left(\frac{3}{2}R - 1\right)\frac{D}{Dt}\mathbf{u}(\mathbf{r}(t),t). \tag{28}$$

This can formally be obtained by expanding Eq.(5) around its $\mathrm{St} = 0$ solution: $\dot{\mathbf{r}}(t) = \mathbf{u} + W\mathbf{n}$ up to first order in St. The advantage of this equation becomes clear when tracing particles backwards in time.

Finding a localised source of particle release is often of central relevance. Such a source-inversion problem appears, for example, in locating a source of air-transported contaminant particles. The approach based on the time-reversed integration of Eq. (5) leads to an unavoidable numerical instability due to an exponential growth of the type $\exp(t/\mathrm{St})$. In contrast, the inertial equation (28) is free from this instability. It can be solved easily in reverse time, too, and this procedure provides with good accuracy the initial spatial coordinates of inertial particles. This has been clearly demonstrated in the example of bubbles in the von Kármán flow [50], and

in other realistic cases like e.g., anthrax in the wind field of an urban street canyon [60] and aerosols in the flow of a hurricane [61].

In spite of the repelling centrifugal force for heavy particles, Vilela and Motter showed [62] that aerosols can also be trapped by open flows under certain circumstances. Such aerosol attractors can exist due to a special interplay of two or more vortices.

In the case of neutrally buoyant particles, $R = 2/3$, in an open chaotic flow, the effect of inertia is to cause a dispersion of particles around the fractal structure of the unstable chaotic set which exists for perfect tracers [63]. Since the introduction of inertia enlarges the phase space to a 4-dimensional manifold (for 2D flows), it is no surprise that the properties of the motion on the 2-dimensional projection to the configuration space are different from its non-inertial counterpart. The important point is that, for small Stokes numbers, the distribution of long-lived inertial particles can still be understood in terms of the simple chaotic set of the non-inertial dynamics. In [63], an expression is derived for the dispersion of particles around the inertia-less chaotic set, which agrees well with numerical simulations. The main result of this work is that inertia causes the fractal structure of the chaotic set to be lost in the configuration space, so that below a certain scale determined by the Stokes number, the spatial distribution becomes smooth. In the slow manifold approach of Haller and coworkers, the slow dynamics coincides with that of infinitesimally small ideal tracers; for $W = 0$, they find that $\dot{\mathbf{r}}(t) = \mathbf{u}$, as can also be seen from (28). Therefore the particle dynamics should synchronise with Lagrangian tracer motion. It has been shown [51], however, that the slow manifold has domains that repel nearby trajectories, which explains the numerical findings of [63].

6 Coagulation and Fragmentation of Finite-Size Particles

So far we have only discussed the motion of passive finite-size particles which are carried by the flow but do not interact with each other. But there is an increasing interest in the investigation of the dynamics of active finite-size particles. The active processes taken into account can be of different natures depending on the context.

In ecology these active finite-size particles, usually aerosols heavier than the fluid, can be plankton species in a limnic or marine environment where plankton populations change their number due to growth and death. Additionally competition and predator-prey interactions influence their dynamics. Particles of different sizes gather along different attractors, as explained in the previous Sections, and therefore different species are expected to occupy different niches, which promotes the coexistence of competitors.

In chemical reactions the active finite-size particles are often bubbles (lighter than the fluid) containing catalysts which mediate particular chemical reactions.

The example which is discussed here in more detail concerns the process of coagulation and fragmentation of finite-size particles in the presence of gravity, which plays an important role in cloud physics [64], marine snow and sediment dynamics

[65, 66], engineering [67], planet formation [68] as well as wastewater treatment [69]. In all the cases mentioned above particles are assumed to interact very rarely, i.e., a kind of "dilute gas" assumption is used.

Coagulation and fragmentation are two processes which influence the size of the particles. Coagulation can happen when two particles collide and form a larger coagulate due to some adhesive forces [64]. Fragmentation is the break-up of a large aggregate into a few smaller ones due to shear forces in the fluid. The dynamics of a system including coagulation and fragmentation is more complicated than the pure advection of finite-size particles. The difficulty lies in the different sizes of the coagulates being advected by the fluid. Because of the dilute approximation, we can assume that the motion of all particles follows the Maxey-Riley equation (5). Coagulates of different sizes correspond to different parameters, namely different Stokes number St and settling velocity W in these equations. Instead of one dynamical system one has to deal with a set of dynamical systems, each of them corresponding to a certain coagulate size. Moreover, the number of coagulates in each dynamical system is changing continuously due to coagulation and fragmentation.

Coagulates of different size converge to different attractors [70] which can be either fixed points, periodic motions, quasiperiodic motions on tori or chaotic attractors (cf. Fig. 17). Among the latter we find a variety of different forms from very localised ones up to space filling attractors where coagulates are distributed over the whole configuration space. As a consequence coagulates of different size are located in different parts of the configuration space in the long-term limit. Hence, a system containing coagulates of different sizes is characterised by an overlay of different attractors possessing possibly different dynamical properties. However, since coagulation and fragmentation happen usually on much smaller time scales than convergence to the attractor, the overall dynamics is in general transient and only a blurred structure of the attractors will be observable.

The same arguments apply to systems where the finite-size particles are bubbles instead of aerosols. Bubbles and aerosols of the same size will occupy different regions in configuration space and exhibit a different kind of dynamics. Therefore, the dynamics of bubbles and aerosols as well as their changes (bifurcations) with

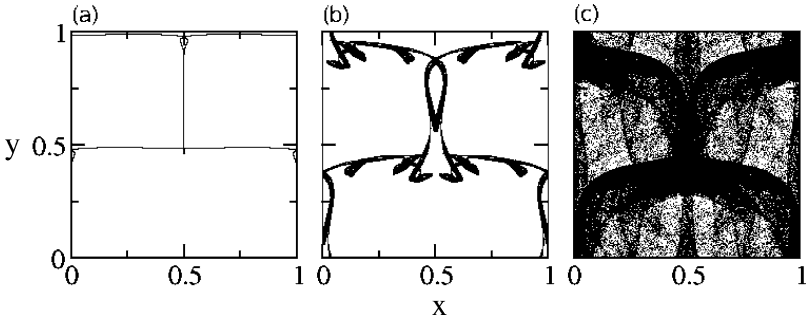


Fig. 17 Attractors for 3 different size classes in a convection flow (9) (cf. Sect. 3.2 for the flow parameters). (**a**) $\mathrm{St}^{-1} = 7.0$, (**b**) $\mathrm{St}^{-1} = 2.778$, (**c**) $\mathrm{St}^{-1} = 2.253$ and $W = 0.4/A$ in all cases

respect to variations of their size are completely different even in the case where their sizes are identical [70]. Finally it is important to note that the form of the attractors depends crucially on the flow [71, 62, 72].

Let us assume that there is a smallest particle in the system which cannot be fragmented. We call it primary particle with radius a_1, mass m_1, Stokes number St_1 and settling velocity W_1. To distinguish the different sizes of the coagulates it is convenient to introduce a size class index α corresponding to the number of primary particles that make up the coagulate. The radius of the coagulate is then $a_\alpha = \alpha^{1/3} a_1$, its Stokes number can be expressed as $\mathrm{St}_\alpha = \alpha^{2/3}\mathrm{St}_1$, and its settling velocity is $W_\alpha = \alpha^{2/3} W_1$; here we see that W/St is independent of α, as discussed at the end of Sect. 2.1. To derive these relations we assume that the coagulates are spherical particles with radius a_α. This assumption applies well to raindrops, while it is a crude approximation for marine aggregates. Marine aggregates are composites of an inorganic kernel like silt or clay with organisms like algae and bacteria attached to it which make up a fractal structure for the whole aggregate. Hence the shape of marine aggregates is more a fractal object than a spherical one.

Coagulation results from a collision of two particles of radii a_i and a_j forming a coagulate of radius $a_{\mathrm{new}}^3 = a_i^3 + a_j^3$ owing to mass conservation. This implies that the size class index of the new coagulate follows from $\alpha_{\mathrm{new}} = \alpha_i + \alpha_j$ which enters $\mathrm{St}_{\mathrm{new}}$ and W_{new}. The velocity of the newly formed coagulate is determined by momentum conservation.

The mechanisms of fragmentation are more complex. While raindrops break apart when they reach a certain maximum size, marine aggregates split up owing to shear forces in the fluid. In the latter case there exists a critical shear force which has to be larger than the intrinsic binding forces of the coagulate to lead to fragmentation. The strength of these binding forces can be measured in terms of a parameter called coagulate strength. For large enough shear forces which overcome the critical shear force a coagulate splits into two smaller ones (determined by a splitting rule, see below). Their velocities are equal owing to momentum conservation and their location is assumed to be directly neighbouring to each other in a random orientation. If the binding force of one of the new smaller coagulates is again smaller than the critical shear force, then an additional break-up takes place. This way more than two new coagulates can result from a single fragmentation event.

Implementing coagulation and fragmentation in the way described above the dynamics of the system can be simulated for different fluid flows. For all subsequent figures concerning coagulation and fragmentation the convection flow explained in Sect. 3.2 is used. Similar results can be obtained for the sine shear flow [72]. In any case one obtains a steady size distribution of coagulates where coagulation and fragmentation balance each other. The shape of the asymptotic size distribution depends on several factors.

Firstly, the size distribution depends on the fluid flow. Secondly, the shape of the size distribution is crucially dependent on the mechanism of fragmentation. While the realization of coagulation is rather straightforward and does not influence the shape of the size distribution, fragmentation needs a more detailed knowledge about the break-up processes to be taken into account. Two different properties of the

fragmentation process are important to consider: (i) what is the critical shear force (splitting condition), how it depends on the size of the coagulate and on the intrinsic binding strength or coagulate strength and, (ii) what is the size distribution of the two fragments created (splitting rule).

To quantify differences in the size distributions depending on various parameters characterising either the flow or the active processes it is convenient to define the average size class index $\langle \alpha(t) \rangle = \sum_\alpha \alpha N_\alpha(t)/N(t)$, where $N_\alpha(t)$ denotes the number of coagulates in size class α, while $N(t)$ is the total number of coagulates in the system. Note that the total number of coagulates always changes in time owing to coagulation and fragmentation. As time evolves $\langle \alpha(t) \rangle$ is found to always reach a limiting value α_∞ though still fluctuating. This implies that a steady size distribution sets in, which is unique, i. e., one finds the same distribution for almost all initial conditions.

Let us now discuss the dependence of the size distribution on the details of the fragmentation process. Unfortunately there are only a few experimental studies devoted to the fragmentation process [73, 74]. Due to this limited knowledge about the details of fragmentation an inverse modelling process has to be considered of making several assumptions about fragmentation and asking which of those yield size distributions which are qualitatively in agreement with observations. In this way one is able to find indications for the most probable fragmentation mechanism.

Fragmentation is modelled by two different mechanisms which occur simultaneously: On the one hand shear forces in the fluid lead to *shear fragmentation* as is typical for marine aggregates and, on the other hand, *size limiting fragmentation* splits all coagulates of a predefined maximum size as is typical for rain drops. The latter break-up process is assumed to be present in shear fragmentation as well. It is known that larger coagulates are more fragile and, hence, fragment more easily. In shear fragmentation the following ansatz for the critical velocity difference Δu_c across a coagulate of size a_α is made:

$$\frac{\Delta u_c}{a_\alpha} \sim \gamma \alpha^{-1/3} \tag{29}$$

where γ is the coagulate strength: a change of this parameter corresponds to considering different types of coagulates. This ansatz is also supported by Taylor [75] and Delichatsios [76] who derived an expression for the critical velocity gradient for the break-up of spherical liquid drops depending on surface tension and viscosity of the drop.

Using Eq. (29) the average size class index is essentially determined by the coagulate strength γ measuring the intrinsic binding strength of the coagulate. With increasing γ the limiting average size class index α_∞ has been found to grow in simulations [72, 77] as

$$\alpha_\infty \sim \gamma^{1/3} \tag{30}$$

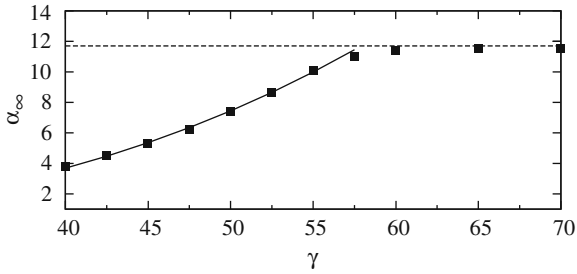


Fig. 18 Asymptotic average size class index α_∞ vs. coagulate strength γ for $\mathrm{St}_1^{-1} = 55.0$, $W_1 = 3.2\mathrm{St}_1$, $\alpha_{\max} = 30$, $a_{\max} = 5 \times 10^{-5}$

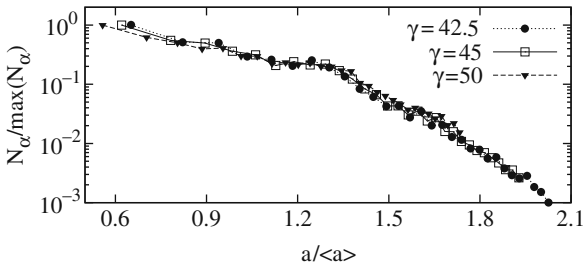


Fig. 19 Normalised size distributions for different coagulate strengths vs. a_α/α_∞ for $\mathrm{St}_1^{-1} = 7.0$, $W_1 = 0.4\mathrm{St}_1$. We started with an initial condition of 10^6 particles with radius $a_1 = 1.5 \times 10^{-5}$

until it reaches (at larger values of γ) a constant finite value (Fig. 18). In the latter interval of γ fragmentation is exclusively dominated by size limiting fragmentation since the binding forces are too strong to allow for shear fragmentation.

In a certain intermediate range of γ where shear fragmentation dominates, all size distributions collapse onto a master curve as shown in Fig. 19 if the steady state size distribution is represented in a normalised form according to $N_\alpha / \max(N_\alpha)$ vs. a_α/α_∞.

As already mentioned the splitting rule determining the size of the two fragments after breaking apart is another important detail of the fragmentation process which can possess several forms. Three different splitting rules have been taken into account in numerical experiments to investigate their impact on the steady size distribution of coagulates: *Uniform splitting* describes a break-up where the size of the first fragment is chosen randomly from a uniform distribution between the smallest size class index 1 and the size class index of the coagulate before splitting α_{old}. For *large scale splitting* the two fragments are of almost equal size, which is expressed mathematically by $\alpha_1 = \alpha_{\mathrm{old}}/2 - |\xi|$ where ξ is a random number (rounded towards the nearest integer) from a normal distribution with zero mean and standard deviation 1, that is cut off at $\pm(\alpha_{\mathrm{old}} - 1)$. *Erosion* corresponds to a break-up where the size of the first fragment is much smaller than the other one, so that $\alpha_1 = |\xi|$ with ξ defined as for large scale splitting.

Numerical simulations show that the shape of the steady size distribution depends crucially on the applied splitting rule. While for erosion and uniform splitting the smallest size class contains most of the coagulates, large scale splitting yields a size distribution with a pronounced maximum at rather small size classes and an exponential decay towards larger size classes (Fig. 20). Particularly this exponential tail

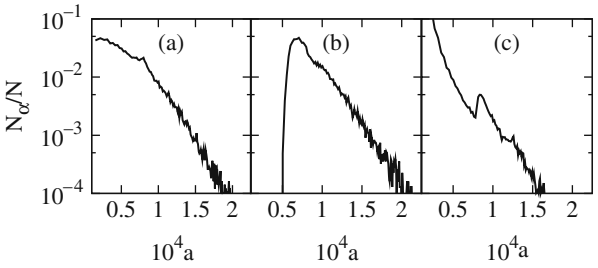


Fig. 20 Size distributions for different splitting rules (**a**) uniform splitting, (**b**) large scale splitting, (**c**) erosion. The parameters are the same as in Fig. 19, $\gamma = 80$

corresponds rather well to observations of size distributions of marine aggregates in tidal areas [78]. This can be interpreted as a strong hint that large scale splitting might be the dominant fragmentation rule for marine aggregates in coastal waters.

The shape of the size distribution is also dependent on the velocity field of the flow. As long as the attractors for all occurring size classes are space-filling, there are no large qualitative differences between the limiting size distributions. Quantitatively, the differences are due to the shear induced by the flow field. However, if the attractors appearing for different flows are not all space filling but differ essentially with respect to their extension in the configuration space (localised vs. space filling) or dynamically (periodic and quasiperiodic vs. chaotic) then large differences in the steady state size distributions may be found.

7 Future Directions

The ubiquity and relevance of suspended particles in time-dependent flows makes it very important to understand properly the dynamics of finite-size particle advection in chaotic flows. This area is in active development, and the subject is the source of many tantalising questions to scientists. We list below what we think are some promising directions of this field.

Active inertial flows. The existing theory of active flows (that is, the dynamics of chemical reactions or other active processes taking place in a flow) is currently formulated mainly for passive tracers [79]. However, this theory only assumes the existence of fractal spatial distribution of the particles in the flow, and exponential contraction towards these fractal filaments. Generally, in case of inertial particles, as discussed extensively in this paper, the fractal structures are present in a phase space which is higher dimensional than the configuration space. If the projection of the phase space to the configuration space shows a fractal distribution of particles, we can expect that the theory developed for passive tracers holds for inertial particles as well. A major difference, however, is that particles of different size or inertia are expected

to occupy slightly different fractal sets. For example, in the case of species coexistence, a recent work shows that inertia can have a dramatic effect on the population dynamics [80], indicating that it is very important in general active processes. The coagulation and fragmentation process of Sect. 6 is another example of an active process for which the particles' inertia is essential.

Non-spherical particles. Up to now most of the work has been devoted to particles of spherical shape; Eq. (2) is only valid for spherical particles. This approximation works well in case of the formation of raindrops, but is only a rather crude approximation in many other applications like e.g. marine aggregates, which have a fractal-like structure. It can be expected that the non-spherical shape of the particles has a large influence on the dynamics of the particles [81]. As an example the aggregation theory of dust particles in planet formation has been discussed by Wilkinson et al. [68]. While fractal-like particles have been investigated in the context of a mean field approach to describe aggregation and fragmentation [82, 83], the consideration of such non-spherical particles in chaotic flows is a topic of current research.

Hydrodynamical interactions between particles. By moving in the flow, particles modify the velocity field in their vicinity, and this change may in its turn affect the motion of another nearby particle. In this way, a hydrodynamical interaction between particles is created which amounts to an inter-particle force. This effect is usually neglected, but may become very important in high particle concentrations. A particularly important question in this context is how the fractal particle distributions created by chaotic advection are affected by the inter-particle interactions [84].

Non-rigid particles. The Maxey-Riley equation assumes a rigid particle, but there are many important cases where the "particle" is non-rigid. An example is raindrops, or any other liquid droplets within a fluid flow [85]. The fact that their shape is variable and depends in particular on the strain makes their treatment challenging, but important.

Neglected terms in the Maxey-Riley equations. In numerical or analytical investigations of finite-size particles in chaotic flows, the history term and the Faxén terms of the Maxey-Riley equation are almost universally neglected by researchers. It is important to have a more rigorous treatment of these terms [86], and to know more precisely under what conditions one can neglect them, and if there are particular flows for which these terms can be important.

We finish with a brief discussion on the case of particles advected by turbulent flows. We have focused throughout this work on the case of non-turbulent (though chaotic) flows. There is a vast literature on effects of inertia in fully developed turbulence, an important subject for many areas. Today it is possible to investigate the particle properties in turbulence at high spatial and temporal resolution. By the appearance of advanced experimental techniques [87, 88] a direct comparison of experiments and numerical simulations is available. The smallest scales in turbulence are given by the Kolmogorov length $\eta = (\nu^3/\epsilon)^{1/4}$ and the Kolmogorov

time $\tau_\eta = (\nu/\epsilon)^{1/2}$, where ϵ denotes the energy dissipation rate [88]. It is natural to define a Stokes number St_η as the ratio of the Stokesian relaxation time $\tau = 2a^2/(9R\nu)$ following from Eq. (2) for particles of size a to τ_η: $\mathrm{St}_\eta = \tau/\tau_\eta$. Note that this Stokes number differs from the one used throughout the paper which is $\mathrm{St} = \tau U/L$ (as can be seen from (6)), i.e., the ratio of τ to the large scale hydrodynamical time. The particle dynamics in turbulence depends on the scale of observation.

On small scales, below the Kolmogorov lengths (in the so-called dissipative range), the flow is smooth, and viscosity dominates. The overall situation is similar to what has been discussed in the bulk of the paper: particles tend to accumulate on chaotic attractors (projected to the space of the flow), and show fractal patterns. The characteristic dimension (e.g., the Kaplan-Yorke dimension [17]) starts to deviate (quadratically) from the dimension n of the flow as St_η takes on small but positive values [89]. The dimension then reaches a minimum value at a Stokes number $\mathrm{St}_{\eta 1}$ of order one, corresponding to a strongest clustering, also called preferential concentration, with a clean fractal structure. A further increase of the Stokes number leads to an increase of dimension, which reaches again the value of n at some $\mathrm{St}_{\eta 2} > \mathrm{St}_{\eta 1}$, still of order one. Beyond this Stokes number, the attractor dimension is larger than n in the full phase space, and its projection on the fluid is space-filling. This scenario appears to be independent of the Reynolds number of the flow [90] but depends smoothly on the density ratio R [91]. An important observation is that the velocity of the particles as a function of the spatial coordinate might be multivalued. Locations where the multivaluedness starts to develop are called caustics [89, 92, 93] and their existence has serious consequences for collision rates of finite-size particles.

Beyond the Kolmogorov length (in the inertial range), the particle distribution is no longer scale invariant, but it is characterised by voids spanning all scales. Heavy particles have been found to cluster where the acceleration is large, i.e., where pressure gradients dominate [89]. Light particles prefer to stay in regions of the flow characterised by rotation [91]. In the inertial range, preferential concentration appears thus to coincide with regions of certain Eulerian characteristics.

The research of the Lagrangian properties of particles in turbulence is rapidly growing (see the review of [88]), and interesting new insights are likely to emerge. Of particular interest is the behaviour of inertial particles in non-ideal (e.g., not fully developed) turbulence. When a large scale flow is superimposed on ideal turbulence, we expect that the particle dynamics on this scale becomes again similar in nature to that on small scales, and the ideas worked out in the bulk of the paper are then directly applicable.

Acknowledgement U.F. would like to thank Jens C. Zahnow for preparing the figures for Sect. 6. J.H.E.C. and O.P. would like to acknowledge the contributions their colleagues Armando Babiano, Mario Feingold, Marcelo Magnasco, Antonello Provenzale and Idan Tuval have made to their understanding of this subject.

References

1. M. R. Maxey and J. J. Riley. *Equation of motion for a small rigid sphere in a nonuniform flow*. *Phys. Fluids*, **26**, 883 (1983).
2. T. R. Auton, F. C. R. Hunt, and M. Prud'homme. *The force exerted on a body in inviscid unsteady non-uniform ratational flow*. *J. Fluid. Mech.*, **197**, 241 (1988).
3. G. I. Taylor. *The forces on a body placed in a curved or converging stream of fluid*. *Proc. R. Soc. London A*, **120**, 260 (1928).
4. H. Aref. *Stirring by chaotic advection*. *J. Fluid Mech.*, **143**, 1 (1984).
5. J. M. Ottino. *The Kinematics of Mixing: Stretching, Chaos and Transport*. Cambridge University Press, Cambridge (1989).
6. H. Aref. *Chaos Applied to Fluid Mixing*. Pergamon Press, Oxford (1995).
7. E. E. Michaelides. *The transient equation of motion for particles, bubbles, and droplets*. *J.Fluids Eng.*, **119**, 233 (1997).
8. J. Magnudet and I. Eames. *The motion of high-Reynolds-number bubbles in inhomogeneous flows*. *Annu. Rev. Fluid. Mech.*, **32**, 659 (2000).
9. E. E. Michaelides. *Particles, Bubbles and Drops: Their Motion, Heat and Mass Transfer*. World Scientific, Singapore (2006).
10. A. B. Basset. *On the motion of a sphere in a viscous liquid*. *Phil. Trans. Roy. Soc*., **179**, 43 (1888).
11. J. Boussinesq. *Sur la résistance qu'oppose un fluide indéfini en repos, sans pesanteur, au mouvement varié d'une sphère solide qu'il mouille sur toute sa surface, quand les vitesses restent bien continues et assez faibles pour que leurs carrés et produits soient négligeables*. *C. R. Acad. Sci. Paris*, **100**, 935 (1885).
12. M. R. Maxey. *The gravitational settling of aerosol-particles in homogeneous turbulence and random flow-fields*. *J. Fluid. Mech.*, **174**, 441 (1987).
13. J. J. B. Biemond et al. *Onset of chaotic advection in open flows*. *Phys. Rev. E*, **78**, 016317 (2008).
14. M. Van Dyke. *An Album of Fluid Motion*. Parabolic Press, Stanford (1982).
15. J. C. Sommerer et al. *Experimental evidence for chaotic scattering in a fluid wake*. *Phys. Rev. Lett.*, **77**, 5055 (1996).
16. J. H. E. Cartwright, M. Feingold, and O. Piro. "An introduction to chaotic advection." In H. Chate, E. Villermaux, and J. M. Chomez, editors, *Mixing: Chaos and Turbulence*. Kluwer, Dordrecht, pp. 307–342 (1999).
17. L. D. Landau and E. M. Lifshits. *Fluid Mechanics*. Elsevier, Butterworth-Heinemann, Oxford, UK (2000).
18. G. K. Batchelor. *An Introduction to Fluid Dynamics*. Cambridge University Press, Cambridge (1967).
19. E. Ott. *Chaos in Dynamical Systems*. Cambridge University Press, Cambdridge (2002).
20. T. Tél and M. Gruiz. *Chaotic Dynamics*. Cambridge University Press, Cambridge (2006).
21. S. Chadrasekhar. *Hydrodynamic and Hydromagnetic Stability*. Oxford University Press, Oxford (1961).
22. S. Wiggins. *Global Bifurcations and Chaos*. Springer, New York (1988).
23. T. Nishikawa et al. *Finite-size effects on active chaotic advection*. *Phys. Rev. E*, **65**, 026216 (2002).
24. P. K. Kundu. *Fluid Mechanics*. Academic Press, New York (1990).
25. C. Jung, T. Tél, and E. Ziemniak. *Application of scattering chaos to particle transport in a hydrodynamical flow*. *Chaos*, **3**, 555 (1993).
26. E. Ziemniak, C. Jung, and T. Tél. *Tracer dynamics in open hydrodynamical flows as chaotic scattering*. *Phys. D*, **79**, 424 (1994).
27. C. Jung and E. Ziemniak. *Hamiltonian scattering chaos in a hydrodynamical system*. *J. Phys. A*, **25**, 3929 (1992).

28. Á. Péntek, Z. Toroczkai, T. Tél, C. Grebogi, and Y. Yorke. *Fractal boundaries in open hydrodynamical flows – signatures of chaotic saddles*. *Phys. Rev. E*, 51, 4076 (1995).
29. M. A. Sanjuán, J. Kennedy, E. Ott, and J. A. Yorke. *Indecomposable continua and the characterization of strange sets in nonlinear dynamics*. *Phys. Rev. Lett.*, **78**, 1892 (1997).
30. M. A. F. Sanjuan, J. A. Kennedy, C. Grebogi, and J. A. Yorke. *Indecomposable continua in dynamical systems with noise: Fluid flow past an array of cylinders*. *Chaos*, **7** 125 (1997).
31. I. Scheuring, T. Czárán, P. Szabó, G. Károlyi, and Z. Toroczkai. *Spatial models of prebiotic evolution: soup before pizza? Orig. Life Evol. Biosph.*, **33**, 319 (2003).
32. M. Sandulescu, E. Hernandez-Garcia, C. López, and U. Feudel. *Kinematic studies of transport across an island wake, with application to the Canary islands*. *Tellus A*, **58**, 605 (2007).
33. M. Sandulescu, C. E. López, E. Hernandez-Garcia, and U. Feudel. *Plankton blooms in vortices: the role of biological and hydrodynamic timescales*. *Nonlinear Proc. Geophys.*, **14**, 1 (2007).
34. O. A. Druzhinin and L. A. Ostrovsky. *The influence of Basset force on particle dynamics in 2-dimensional flows*. *Phys. D*, **76**, 34 (1994).
35. P. Tanga and A. Provenzale. *Dynamics of advected tracers with varying buoyancy*. *Phys. D*, **76**, 202 (1994).
36. A. N. Yannacopoulos, G. Rowlands, and G. P. King. *Influence of particle inertia and Basset force on tracer dynamics: Analytic results in the small-inertia limit*. *Phys. Rev. E*, **55**, 4148 (1997).
37. J. R. Angilella. *Asymptotic analysis of chaotic particle sedimentation and trapping in the vicinity of a vertical upward streamline*. *Phys. Fluids*, **19**, 073302 (2007).
38. J. R. Angilella. *Chaotic particle sedimentation in a rotating flow with time-periodic strength*. *Phys. Rev. E*, **78**, 066310 (2008).
39. F. Candelier, J. R. Angilella, and M. Souhar. *On the effect of the Boussinesq-Basset force on the radial migration of a Stokes particle in a vortex*. *Phys. Fluids*, **16**, 1765 (2004).
40. F. Candelier and J. R. Angilella. *Analytical investigation of the combined effect of fluid inertia and unsteadiness on low-Re particle centrifugation*. *Phys. Rev. E*, **73**, 047301 (2006).
41. A. Babiano, J. H. E. Cartwright, O. Piro, and A. Provenzale. *Dynamics of a small neutrally buoyant sphere in a fluid and targeting in Hamiltonian systems*. *Phys. Rev. Lett.,* **84**, 5764–5767 (2000).
42. A. Crisanti, M. Falcioni, A. Provenzale, and A. Vulpiani. *Passive advection of particles denser than the surrounding fluid*. *Phys. Lett. A*, **150**, 79 (1990).
43. A. Okubo. *Horizontal dispersion of floatable particles in vicinity of velocity singularities such as convergences*. *Deep-Sea Res.*, **17**, 445 (1970).
44. J. B. Weiss. *The dynamics of enstrophy transfer in 2-dimensional hydrodynamics*. *Phys. D*, **48**, 273 (1991).
45. D. Elhmaidi, A. Provenzale, and A. Babiano. *Elementary topology of 2-dimensional turbulence from a Lagrangian viewpoint and single-particle dispersion*. *J. Fluid Mech.*, **257**, 533 (1993).
46. C. Basdevant and T. Philipovich. *On the validity of the Weiss criterion in 2-dimensional turbulence*. *Phys. D*, **73**, 17 (1994).
47. B. L. Hua and P. Klein. *An exact criterion for the stirring properties of nearly two-dimensional turbulence*. *Phys. D*, **113**, 98 (1998).
48. B. Protas, A. Babiano, and N. K.-R. Kevlahan. *On geometrical alignment properties of two-dimensional forced turbulence*. *Phys. D*, **128**, 169 (1999).
49. A. Provenzale. *Transport by coherent barotropic vortices*. *Annu. Rev. Fluid Mech.*, **31**, 55 (1999).
50. G. Haller and T. Sapsis. *Where do inertial particles go in fluid flows? Phys. D*, **237**, 573 (2008).
51. T. Sapsis and G. Haller. *Instabilities in the dynamics of neutrally buoyant particles*. *Phys. Fluids*, **20**, 017102 (2008).
52. J. H. E. Cartwright, M. O. Magnasco, and O. Piro. *Bailout embeddings, targeting of invariant tori, and the control of Hamiltonian chaos*. *Phys. Rev. E*, **65**, 045203(R) (2002); J. H. E.

Cartwright, M. O. Magnasco, O. Piro, and I. Tuval. *Bailout embeddings and neutrally buoyant particles in three-dimensional flows*. *Phys. Rev. Lett.*, **89**, 264501 (2002); J. H. E. Cartwright, M. O. Magnasco, O. Piro, and I. Tuval. *Noise-induced order out of chaos by bailout embedding*. *Fluctuation Noise Lett.*, **2**, 161 (2002); J. H. E. Cartwright, M. O. Magnasco, O. Piro, and I. Tuval. *Bubbling and on-off intermittency in bailout embeddings*. *Phys. Rev. E*, **68**, 016217 (2003).

53. J. H. E. Cartwright, M. O. Magnasco, O. Piro, and I. Tuval. *Noise- and inertia-induced inhomogeneity in the distribution of small particles in fluid flows*. *Chaos*, **12**, 489 (2002).
54. V. I. Arnold. *Sur la topologie des écoulements stationnaires des fluides parfaits*. *C. R. Acad. Sci. Paris A*, **261**, 17 (1965); M. Hénon. *Sur la topologie des lignes de courant dans un cas particulier*. *C. R. Acad. Sci. Paris A*, **262**, 312 (1966).
55. T. Dombre et al. *Chaotic streamlines in the ABC flows*. *J. Fluid Mech.*, **167**, 353 (1986).
56. I. J. Benczik, Z. Toroczkai, and T. Tél. *Selective sensitivity of open chaotic flows on inertial tracer advection: Catching particles with a stick*. *Phys. Rev. Lett.*, **89**, 164501 (2002).
57. I. Benczik, T. Toroczkai, and T. Tél. *Advection of finite-size particles in open flows*. *Phys. Rev. E*, **67**, 036303 (2003).
58. Y. Do and Y.-C. Lai. *Superpersistent chaotic transients in physical space: advective dynamics of inertial particles in open chaotic flows under noise*. *Phys. Rev. Lett.*, **91**, 224101 (2003).
59. G. Haller and T. Sapsis. *Localized instability and attraction along invariant manifolds*. (2009).
60. W. Tang, G. Haller, J.-J. Baik, and Y.-H. Tyu. *Locating an atmospheric contamination source using slow manifolds*. *Phys. Fluids*, **21**, 043302 (2009).
61. T. Sapsis and G. Haller. Inertial particle dynamics in a hurricane. *J. Atmos. Sci.*, **66**, 2481–2492 (2009).
62. R. D. Vilela and A. E. Motter. *Can aerosols be trapped in open flows? Phys. Rev. Lett.*, **99**, 264101 (2007).
63. R. D. Vilela, A. P. S. de Moura, and C. Grebogi. *Finite-size effects on open chaotic advection*. *Phys. Rev. E*, **73**, 026302 (2006).
64. H. R. Pruppacher and J. D. Klett. *Microphysics of Clouds and Precipitation*. Kluwer Academic Publishers, Dordrecht (1997).
65. G. A. Jackson. *A model of the formation of marine algal flocs by physical coagulation processes*. *Deep-Sea Res.*, **37**, 1197 (1990).
66. W. McAnally and A. Mehta. *Significance of aggregation of fine sediment particles in their deposition*. *Estuarine Coastal Shelf Sci.*, **54**, 643 (2002).
67. C. Crowe, M. Sommerfeld, and Y. Tsuji. *Multiphase Flows with Particles and Droplets*. CRC Press, New York (1998).
68. M. Wilkinson et al. *Stokes trapping and planet formation*. *Astrophys. J. Suppl. Ser.*, **176**, 484 (2008).
69. J. Zhang and X.-Y. Li. *Modeling particle-size distribution dynamics in a flocculation system*. *AIChE J.*, **49**, 1870 (2003).
70. J. C. Zahnow and U. Feudel. *Moving finite-size particles in a flow: A physical example of pitchfork bifurcations of tori*. *Phys. Rev. E*, **77**, 026215 (2008).
71. T. Nishikawa et al. *Advective coalescence in chaotic flows*. *Phys. Rev. Lett.*, **87**, 038301 (2001).
72. J. C. Zahnow, R. D. Vilela, U. Feudel et al. *Coagulation and fragmentation dynamics of inertial particles*. *Phys. Rev. E*, **80**, 026311 (2009).
73. E. Villermaux. *Fragmentation*. *Annu. Rev. Fluid Mech.*, **39**, 419 (2007).
74. A. Alldredge et al. *The physical strength of marine snow and its implications for particle disaggregation in the ocean*. *Limnol. Oceanogr.*, 35,1415 (1990).
75. G. I. Taylor. *The formation of emulsions in definable fields of flow*. *Proc. Roy. Soc. A*, **146**, 501 (1934).
76. M. Delichatsios. *Model for breakup rate of spherical drops in isotropic turbulent flows*. *Phys. Fluids*, **30**, 622 (1975).
77. J. C. Zahnow et al. *Aggregation and fragmentation dynamics of inertial particles in chaotic flows*. *Phys. Rev. E*, **77**, 055301(R) (2008).

78. M. Lunau et al. *Physical and biogeochemical controls of microaggregate dynamics in a tidally affected coastal ecosystem*. *Limnol. Oceanogr.*, **51**, 847 (2006).
79. Z. Toroczkai, G. Károlyi, A. Péntek, T. Tél, and C. Grebogi. *Advection of active particles in open chaotic flows*. *Phys. Rev. Lett.*, **80**, 500 (1998); T. Tél, A. P. S. de Moura, G. Károlyi, and C. Grebogi. *Chemical and biological activity in open flows: A dynamical system approach*. *Phys. Rep.*, 413, 91 (2005); J. H. E. Cartwright, J. M. Garcia-Ruiz, O.Piro, C. I. Sainz-Diaz, and I. Tuval. *Chiral symmetry breaking during crystallization: An advection-mediated nonlinear auto- catalytic process. Phys. Rev. Lett.* **93**, 035502 (2004); J. H. E. Cartwright, O.Piro, and I. Tuval. *Ostwald ripening, chiral crystallization, and the common-ancestor effect. Phys. Rev. Lett.,* **98**, 165501 (2007).
80. I. J. Benczik, G. Károlyi, I. Scheuring, and T. Tél. *Coexistence of inertial competitors in chaotic flows*. *Chaos*, **16**, 043110 (2006).
81. H. Shin and M. R. Maxey. *Chaotic motion of nonspherical particles settling in a cellular flow field*. *Phys. Rev. E*, **56**, 5431 (1997).
82. C. Kranenburg. *The fractal structure of cohesive sediment aggregates*. *Estuarine Coastal Shelf Sci.*, **39**, 541 (1994).
83. F. Maggi, F. Mietta, and J. C. Winterwerp. *Effect of variable fractal dimension on the floc size distribution of suspended cohesive sediment*. *J. Hydrol.*, **343**, 43 (2007).
84. R. Medrano, A. Moura, T. Tel, I. Caldas, and C. Grebogi. *Finite-size particles, advection, and chaos: A collective phenomenon of intermittent bursting*. *Phys. Rev. E,* **78**, 56206 (2008).
85. M. Tjahjadi and J. M. Ottino. *Stretching and breakup of droplets in chaotic flows*. *J. Fluid Mech.,* **232**, 191 (1991); J. M. Ottino. *Unity and diversity in mixing: Stretching, diffusion, breakup, and aggregation in chaotic flows. Phys. Fluids A*, **3**, 1417 (1991).
86. N. Mordant and J. -F. Pinton. *Velocity measurement of a settling sphere*. *Eur. Phys. J.* **B 18**, 343–352 (2000).
87. N. T. Ouellette, P. J. O'Malley, and J. P. Gollub. *Transport of finite-sized particles in chaotic flow*. *Phys. Rev. Lett.*, **101**, 174504 (2008).
88. F. Toschi and E. Bodenschatz. *Lagrangian properties of particles in turbulence*. *Annu. Rev. Fluid. Mech.*, **41**, 375 (2009).
89. J. Bec. *Fractal clustering of inertial particles in random flows*. *Phys. Fluids*, **15**, L81 (2003).
90. J. Bec et al. *Heavy particle concentration in turbulence at dissipative and inertial scales*. *Phys. Rev. Lett.*, **98**, 084502 (2008).
91. E. Calzavarini, M. Kerscher, D. Lohse, and F. Toschi. *Dimensionality and morphology of particle and bubble clusters in turbulent flow*. *J. Fluid Mech.*, **607**, 13 (2008).
92. G. Falkovich, A. Fouxon, and M. G. Stepanov. *Acceleration of rain initiation by cloud turbulence*. *Nature*, **419**, 151 (2002).
93. M. Wilkinson and B. Mehlig. *Caustics in turbulent aerosols*. *Europhys. Lett.*, **71**, 186 (2005).
94. H. Faxén. *Der Widerstand gegen die Bewegung einer starren Kugel in einer zähen Flüssigkeit, die zwischen zwei parallelen ebenen Wänden eingeschlossen ist*. *Ann. Phys.*, **4**, 89 (1922).

Langevin Equation for Slow Degrees of Freedom of Hamiltonian Systems

R.S. MacKay

Abstract A way is sketched to derive a Langevin equation for the slow degrees of freedom of a Hamiltonian system whose fast ones are mixing Anosov. It uses the Anosov-Kasuga adiabatic invariant, martingale theory, Ruelle's formula for weakly non-autonomous SRB measures, and large deviation theory.

1 Introduction

Model reduction is a central theme in science. In particular, it is common to propose to replace "inessential" details of parts of dynamics by noise. This paper addresses the question of to what extent such reduction may be justified if one starts from a deterministic Hamiltonian systems, the agreed foundation for all classical mechanics.

Suppose a Hamiltonian system consists of some slow degrees of freedom coupled to a large number of fast chaotic degrees of freedom. For example, consider the conformation change degrees of freedom of a biomolecule coupled to its vibrations and movement of water molecules. It is standard to model the slow degrees of freedom by a Langevin equation, that is a stochastic differential equation where the effects of the fast degrees of freedom have been replaced by an effective Hamiltonian system, damping and noise.

Despite being in use now for 100 years [16], it seems to me there is not yet a satisfactory derivation of a Langevin equation for the slow degrees of freedom. See the reviews [11, 14] for the state of affairs. An example system which has recently resurfaced is the "piston problem" [18].

One precursor is [8] for the case of Hamiltonians which are quadratic in the fast degrees of freedom and subject to a continuum approximation for the distribution

R.S. MacKay (✉)
Mathematics Institute and Centre for Complexity Science, University of Warwick, Coventry CV4 7AL, UK
e-mail: R.S.MacKay@warwick.ac.uk

M. Thiel et al. (eds.), *Nonlinear Dynamics and Chaos: Advances and Perspectives,* Understanding Complex Systems, DOI 10.1007/978-3-642-04629-2_5,

of frequencies of their normal modes. This approach was presented nicely by [31], and treated more throughly in [28], but the crucial assumption (for the analysis) of harmonic fast degrees of freedom seems unrealistic and overly restrictive to me.

A second precursor is the line of investigation of the sequence of papers [25, 30, 3, 13], which depends on assuming ergodicity of the fast dynamics restricted to an energy level for frozen slow degrees of freedom, but I don't think these papers can be considered to make a mathematically complete derivation of a Langevin equation. As a side-remark, ergodicity may be considered unrealistic and overly restrictive, but perhaps true ergodicity is not really required.

This paper seeks to put more mathematical flesh on the ideas of the second approach, under the stronger assumption that the fast dynamics is mixing Anosov on relevant energy levels for frozen slow degrees of freedom. Although this is a yet more restrictive assumption than ergodicity, [9] advocated that it may be reasonable to assume that the dynamics of a large Hamiltonian system act as if mixing Anosov on relevant energy levels (the "chaotic hypothesis") and that results derived under this assumption may apply more widely. It is worth noting that a two-degree of freedom physically relevant mixing Anosov system has been constructed [12] and I believe analogues can be made in higher degrees of freedom.

Under the mixing Anosov assumption, some nice mathematical results can be applied, namely a formula of Ruelle for the first-order effect of weak non-autonomy on the natural measure for a chaotic system [27], and an almost sure invariance principle to approximate the integral of a zero-mean vector of observables for a chaotic system by a multidimensional Brownian walk [23]. With some work I believe that these ingredients can be put together to derive a Langevin equation for the slow degrees of freedom. The paper sketches the main lines of the proposal.

2 Assumptions

1. Suppose (M,ω) is a symplectic manifold of large dimension $2m$ and $H : M \to \mathbb{R}$ is a smooth function (the Hamiltonian). Together they define a Hamiltonian vector field X_H on M, by $\omega(X_H,\xi) = dH(\xi)$ for all tangent vectors ξ. Equivalently, the symplectic form ω defines a Poisson bracket $\{F,G\} = \omega(X_F,X_G)$ on smooth functions $F,G:M \to \mathbb{R}$, and then $dF(X_H) = X_H(F) = \{F,H\}$ for all $F:M \to \mathbb{R}$. Let ϕ_t be the flow of X_H.
2. Suppose N is a symplectic manifold of moderate dimension $2n$, representing the slow degrees of freedom. It will be enough to do local analysis in N, so without loss of generality it can be considered to be a piece of $\mathbb{R}^{2n}$ with local coordinate functions Z_j. Suppose $\pi:M \to N$ is a Poisson map, i.e. for all smooth functions $F,G:N \to \mathbb{R}$, $\{F \circ \pi, G \circ \pi\}_M = \{F,G\}_N \circ \pi$.
3. Suppose for each $Z \in N$, the "fibre" $\pi^{-1}(Z)$ is a symplectic submanifold of M, i.e. the restriction of ω to its tangent space is everywhere non-degenerate. It follows that the restriction H_Z of H to $\pi^{-1}(Z)$ defines a Hamiltonian vector field X_{H_Z} on $\pi^{-1}(Z)$ (the constrained vector field). Furthermore, X_{H_Z} preserves the $2(m-n)$-dimensional volume form

$$\Omega_Z = \omega^{\wedge(m-n)}/(m-n)!$$

on $\pi^{-1}(Z)$, the level sets $(H,\pi)^{-1}(E,Z)$, and the $(2m-2n-1)$-dimensional volume form $\mu_{Z,E}$ on $(H,\pi)^{-1}(E,Z)$ defined by $\mu \wedge dH = \Omega_Z$. Denote its flow by ψ_t.

4. Suppose

$$W_Z(E) := \int_{\{H \le E\}} \Omega_Z$$

is finite for relevant Z,E and differentiable (actually, $W_Z(E)$ finite is not really necessary as long as its derivative with respect to Z and E is definable). Its E-derivative can be written as

$$W_Z'(E) = \int_{H^{-1}(E)} \mu_{Z,E}.$$

Then $\mu_{Z,E}$ induces an invariant probability measure

$$\lambda_{Z,E}(A) = \frac{1}{W_Z'(E)} \int_A \mu_{Z,E}$$

for subsets A on the level set for X_{H_Z}, which Boltzmann called an "ergode" and Gibbs a "microcanonical ensemble". Subscripts (Z,E) on measures μ and λ will often be dropped. For a function $g:\pi^{-1}(Z) \to \mathbb{R}$, its mean with respect to λ will be written as $\lambda(g)$.

5. Suppose

$$V_j = \{Z_j \circ \pi, H\}$$

are slow compared to X_{H_Z}, in a sense to be made more precise in the next assumption.

6. Suppose X_{H_Z} is mixing Anosov on the level sets for relevant E,Z. In particular, suppose the autocorrelation of the deviation $v(s)$ of $V \circ \psi_s$ from its mean decays in a short time $\varepsilon \ll 1$ on the "slow" timescale, that for significant change in Z under the mean of V.
7. Suppose the size of v (in units for the slow timescale) scales like $\sqrt{\delta/\varepsilon}$ for some δ.
8. Suppose the temperature $T > 0$ is significantly higher than a value determined by δ and the symplectic form on N, to be made explicit in Sect. 4.3.
9. Finally, suppose the heat capacity per fast degree of freedom is positive and bounded.

3 Aim

The aim is to show that for ε small the distribution of paths $\pi \circ \phi_t(Y)$ for random Y with respect to λ_{Z_0,E_0} and for t in any order-one time interval, is close to that for the solutions of a system of stochastic ordinary differential equations

$$dZ_i = (J_{ij} - \beta D_{ij})\nabla_j F \, dt + \sigma_{ik} \, dW_k, \quad Z(0) = Z_0, \tag{1}$$

with $\nabla_j F$ a shorthand for $\frac{\partial F}{\partial Z_j}$, J representing the Poisson bracket on N (i.e. $\{F,G\} = \nabla_i F J_{ij} \nabla_j G$), β the inverse temperature for the initial energy, $F{:}N \to \mathbb{R}$ the free energy function at the given temperature,

$$D_{ij} = \int_{-\infty}^{0} \lambda_{Z,E}(v_i(0)v_j(s)) \, ds, \tag{2}$$

W a multidimensional Wiener process, σ any matrix function satisfying the Einstein-Sutherland relation $\sigma\sigma^T = D + D^T$, and the Klimontovich interpretation of the stochastic differential equation.

Definitions of β, F and the Klimontovich interpretation will be recalled at the appropriate points.

Equation (1) with Klimontovich interpretation implies Klein-Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} = -\mathrm{div}\,(\rho(J - \beta D)\nabla F - S\nabla\rho) \tag{3}$$

for the evolution of probability densities ρ with respect to volume $\wedge_j dZ_j$ on N, where $S = (D + D^T)/2$, the symmetric part of D.

4 Strategy

The strategy of proof has six stages. First $\lambda(V)$ is evaluated in terms of the microcanonical free energy. Secondly, the fluctuations $v(t)$ of V from its mean are approximated by a multi-dimensional white noise with covariance $D + D^T$. Thirdly, a correction to the ergode is derived when Z is moving, which in general yields damping. Fourthly, the effect of autonomous rather than externally imposed Z-motion is argued to make no difference, to the order of approximation considered. Fifthly, the ergode is approximated by the monode (canonical ensemble) for large number of degrees of freedom. Lastly, the Klimontovich interpretation is shown to be necessary.

4.1 Zeroth Order Mean Velocity

Anosov and Kasuga (e.g. [19]) proved that $W_Z(E)$ is an adiabatic invariant for X_{H_Z} ergodic on energy levels, with respect to slow external change of Z, i.e. for most trajectories the energy $E(t)$ changes in such a way to keep $W_{Z(t)}(E(t)) \approx w_0$. Define the "microcanonical free energy" $f{:}N \to \mathbb{R}$ for given w_0 by

$$f(Z) = W_Z^{-1}(w_0). \tag{4}$$

The following calculation shows that

$$\lambda(V) = J\nabla f. \tag{5}$$

Firstly, $W_Z(f(Z)) = w_0$, so $\nabla W + W'\nabla f = 0$, i.e.

$$\nabla f = -\frac{1}{W'}\nabla W. \tag{6}$$

Thus $(J\nabla f)_j = \{Z_j,f\} = -\frac{1}{W'}\{Z_j,W\}$. Next, the flow χ_u of $X_{Z_j\circ\pi}$ preserves Ω (because it is Hamiltonian) and the fibration π (because π is Poisson). Thus the change of $W_Z(E) = \int_{\{H\le E\}} \Omega$ from $Z(0)$ to $Z(u)$ along the flow χ_u is the Ω-measure of the band in $\pi^{-1}(Z(u))$ between $H^{-1}(E)$ and $\chi_u((H,Z)^{-1}(E,Z(0)))$. The rate of change of H along the flow χ_u is $\{H,Z_j \circ \pi\}$ and we can write $\Omega = \mu \wedge dH$ in a fibre, so

$$\{W,Z_j\} = -\int_{H^{-1}(E)} \{H,Z_j \circ \pi\}\mu.$$

Finally, $\lambda(V_j) = \frac{1}{W'}\int_{H^{-1}(E)}\{Z_j \circ \pi,H\}\mu$.

Remark 1 A similar calculation with respect to the canonical ensemble $e^{\beta F}e^{-\beta H}\Omega$ on $\pi^{-1}(Z)$ (Boltzmann's "monode") gives mean velocity $J\nabla F$, where the canonical free energy F is defined by

$$e^{-\beta F(Z)} = \int_{\pi^{-1}(Z)} e^{-\beta H}\Omega_Z. \tag{7}$$

Nevertheless, the canonical ensemble is not ergodic (energy is conserved), so it is not clear how to proceed further (unless we just ignore the mixing requirement for the next results).

Anosov's averaging theorem (e.g. [19]) proves that most trajectories follow the mean dynamics closely on short enough timescales. Thus from (5) the zeroth order dynamics of the slow degrees of freedom is Hamiltonian with Hamiltonian function the microcanonical free energy f. Note in particular that the zeroth order dynamics preserves f, which is a restatement of the adiabatic invariance of $W_Z(E)$.

I am interested in capturing effects that go beyond zeroth order, however, in particular on longer timescales where the invariance of $W_Z(E)$ breaks down. Results of [5] address this, but not in a specifically Hamiltonian context.

4.2 Fluctuations

The fluctuations $v(t)$ of $V \circ \psi_t$ from its mean for fixed Z produce an effect approximately equivalent to white noise with covariance matrix $R = \int_{-\infty}^{\infty} C(t)\, dt$, where $C_{ij}(t) = \lambda(v_i(t)v_j(0))$, provided the integral converges. This is well known but derivations vary in level of sophistication.

The simplest version is to let $z(t) = \int_0^t v(s)\, ds$ (I denote it by z rather than Z because this expression does not include the mean velocity of Z nor the fact that the distribution of v changes as Z moves) and prove that

$$\lambda(z_i(t)z_j(t))/t \to R_{ij}$$

as $t \to +\infty$ (Green-Kubo formula, actually discovered by GI Taylor [29]).

Here is a proof. From the definitions of z and C, $\lambda(z_i(t)z_j(t)) = \int_{-t}^{t} (t - |u|)\, C_{ij}(u)\, du$. So

$$\lambda(z_i(t)z_j(t))/t = \int_{-t}^{t} \left(1 - \frac{|u|}{t}\right) C_{ij}(u)\, du.$$

Tackle the positive and negative ranges of u separately. Convergence of the integral for R implies that given $\varepsilon > 0$ there is a t_0 such that $|\int_u^t C(v)\, dv| \le \varepsilon$ for all $t \ge u \ge t_0$. Then for $t \ge t_0$,

$$\int_0^\infty C(u)\, du - \int_0^t \left(1 - \frac{u}{t}\right) C(u)\, du = \int_t^\infty C(u)\, du + \frac{1}{t}\int_0^{t_0} uC(u)\, du + \frac{1}{t}\int_0^t \int_{\max(u,t_0)}^t C(v)\, dv\, du.$$

The first and third terms are each at most ε in absolute value. The second is at most ε as soon as $t \ge \frac{1}{\varepsilon}\int_0^{t_0} uC(u)\, du$. Hence $\int_0^t \left(1 - \frac{u}{t}\right) C(u)\, du \to \int_0^\infty C(u)\, du$ as $t \to \infty$. Similarly for $u < 0$ and hence the result. Note that in contrast to a statement in [10] it is not necessary to assume $\frac{1}{T}\int_{-T}^{T} |t|C(t)\, dt \to 0$ as $T \to \infty$: it follows automatically from $\int_{-\infty}^{\infty} C(t)\, dt < \infty$.

As a corollary this shows the covariance matrix to be non-negative. The covariance matrix R can be written as $D + D^T$, with D as defined in (2). Thus it also follows that the symmetric part S of D is non-negative.

More sophisticated results use the theory of martingales: stochastic processes on vector spaces such that the expectation of the future value given the present one is just the present value. The best is to prove a "vector-valued almost sure invariance

principle", in the sense that the paths $z(T)$ in $\mathbb{R}^{2n}$ are distributed within $O(T^{\delta})$ of the distribution for the corresponding Brownian paths, for some $\delta < 1/2$. This follows from a general result of [23] provided the fast dynamics is sufficiently rapidly mixing (exponential suffices).

To increase the accuracy of the approximation by white noise, it could be a good idea to iteratively improve the choice of fibration π to make C decay as fast as possible, in particular to remove major changes in sign.

4.3 Correction to Ergode

If $Z(t)$ is moved slowly along some path then the natural probability measure on the moving fast system lags slightly behind the instantaneous ergode $\lambda_{Z,E}$ for the given value of w_0. If we assume that $W_Z(E)$ is conserved exactly, the first order difference can be computed by a formula of [27] (extrapolating a bit beyond his hypotheses, but see [7] for a statement of what should suffice).

Ruelle's formula assumes a direction of time, in the sense that the probability distribution for the fast system is assumed to be absolutely continuous along unstable manifolds, whereas one could have asked for it along stable manifolds. Many people regard this as justified by a hypothesis of a low entropy initial condition for the universe [17], but hypotheses on initial conditions seem to me inadequate to explain the direction of causality. I suspect a true explanation lies in quantum gravity: probably a consistent theory of quantum gravity will exhibit two phases, differing in the direction of interaction of radiation and matter. Our patch of spacetime is in one of these phases and we choose to orient time accordingly.

For a slowly varying vector field X_t on a manifold M, which at each time t is mixing Anosov, Ruelle's formula specifies the first order change to the expectation $\langle O(t)\rangle$ of any smooth observable $O{:}M \to \mathbb{R}$ at time t from that for the frozen system X_t:

$$\delta\langle O(t)\rangle = \int_{-\infty}^{t} \langle d(O \circ \psi_{ts})(X_s - X_t)\rangle \, ds, \tag{8}$$

where ψ_{ts} is the flow of X_t from time s to time t. The term X_t can be dropped since $\langle d(O\circ\psi_{ts})X_t\rangle$ is the expectation of the rate of change of a function with respect to an invariant measure, so zero (it was included to make clear that the result is first order in the change in the vector field). In our case, the state space $\pi^{-1}(Z)$ also changes with time so evaluating X_s requires choosing some diffeomorphism between the state spaces at times s and t, but the result does not depend on the choice.

Let us calculate the change in the mean of V due to slow motion of Z. For X_t we use $X_{H_Z(t)}$ and for the ensemble average we use $\lambda_{Z(t),E}$. Any motion of Z can be specified as the result of a (possibly time-dependent) Hamiltonian flow on N, with some Hamiltonian G, so $\dot{Z} = J\nabla G$. The function G lifts to $G \circ \pi$ on M and so induces a fibre-preserving flow χ on M, which we can use to identify points of different fibres. In particular for X_s in Ruelle's formula we can use $\chi_{ts}^{*}X_{H_Z(s)}$, which can be written as $X_{(H\circ\chi_{st})_{Z(t)}}$. Then

$$d(V_j \circ \psi_{ts})X_s = \{V_j \circ \psi_{ts}, H \circ \chi_{st}\}_{Z(t)}, \tag{9}$$

where $\{,\}_Z$ is the Poisson bracket on $\pi^{-1}(Z)$, defined via the restriction of the symplectic form to the fibre. Thus Ruelle's formula gives a time-integral of an energy-level average of a Poisson bracket on a fibre.

Lemma 1 *For symplectic manifold K with volume form Ω, Hamiltonian H, energy level E, normalised energy level volume λ_E and any smooth functions $F,G:K \to \mathbb{R}$ for which the required integrals converge,*

$$\int \{F,G\}\, d\lambda_E = \frac{1}{W'(E)} \frac{\partial}{\partial E} \left(W'(E) \int \{F,H\} G\, d\lambda_E \right). \tag{10}$$

Proof For any smooth functions $F,U:K \to \mathbb{R}$ for which the integral converges,

$$\int \{F,U\}\, d\Omega = \int dF(X_U)\, d\Omega = 0,$$

because it is the integral of the rate of change of F along orbits of X_U with respect to an invariant measure. Apply this to a product $U = GA$ and use Leibniz' rule and antisymmetry for Poisson brackets to deduce that

$$\int A\{F,G\}\, d\Omega = \int \{A,F\} G\, d\Omega. \tag{11}$$

Now take A to be (a sequence of smooth approximations to) $\delta(E - H)$:

$$\int \delta(E-H)\{F,G\}\, d\Omega = \int \{\delta(E-H),F\} G\, d\Omega = -\int \delta'(E-H)\{H,F\} G\, d\Omega, \tag{12}$$

since $\{.,F\}$ is a derivation. The right hand side can be written as

$$\frac{\partial}{\partial E} \int \delta(E - H)\{F,H\} G\, d\Omega.$$

All that remains is to write $\delta(E - H)\, d\Omega = W'(E)\, d\lambda$ on both sides. □

Applying the lemma to (9) produces

$$\delta\langle V_j \rangle = \frac{1}{W'(E)} \frac{\partial}{\partial E} \left(W'(E) \int_{-\infty}^{t} ds \langle \{H, H \circ \chi_{st}\}_Z V_j \circ \psi_{ts} \rangle \right). \tag{13}$$

Now

$$\frac{\partial}{\partial s} H \circ \chi_{st} = -\{H, G \circ \pi\}_M,$$

so for times s out to some decorrelation time, which we supposed to be $\varepsilon \ll 1$ (Assumption 6), we can write to leading order

$$\{H,H \circ \chi_{st}\} = (t-s)\{H,\{H,G \circ \pi\}_M\}_Z.$$

Specialising to $G = Z_k$ gives $\{H,G \circ \pi\}_M = -V_k$. So the integral in (13) becomes

$$\int_{-\infty}^{t} ds\,(t-s)\langle\{H,V_k\}_Z(s)V_j(t)\rangle.$$

To justify this approximation properly requires some hypothesis on the rate of decay of the correlation function of v (probably $\int_{-\infty}^{0} |tC(t)|\ dt < \infty$ suffices). Now $\{H,V_k\}_Z = -\frac{dV_k}{ds}$ along the flow of X_{H_Z}, so integration by parts (with again some assumption about sufficiently rapid convergence of the autocorrelation integral) transforms the integral to

$$-\int_{-\infty}^{t} ds\langle(V_k(s) - \langle V_k\rangle)V_j(t)\rangle = -D_{jk},$$

with D given by (2). Taking G to be an arbitrary linear combination of Z_k yields

$$\delta\langle V(t)\rangle = -\frac{(W'D)'}{W'}J^{-1}\dot{Z}. \tag{14}$$

This result agrees with [3], except that the term "f_1" which they chose to neglect does not arise here. Perhaps I lost it through making a constant $\dot{Z}$ approximation, or perhaps it really gives nothing to leading order.

The derivation requires $\dot{Z}$ to be small, since Ruelle's formula is first order. This is achieved for small enough ε, because the relevant notion of smallness of $\dot{Z}$ is on the fast timescale.

For $k = m - n$ large, one can expect W' to vary much more rapidly than D with energy E. In particular, under the assumption of Sect. 4.5 to come, $W'(E) \sim e^{ks(E/k)}$, where $s(\epsilon)$ is the entropy per fast degree of freedom as a function of the energy ϵ per fast degree of freedom. Then (14) can be approximated by $-\beta DJ^{-1}\dot{Z}$ with

$$\beta = (\log W')', \tag{15}$$

the inverse (canonical) temperature (in distinction to the inverse microcanonical temperature $(\log W)'$, though they should be virtually the same for large k). Thus we obtain

$$\dot{Z} = J\nabla f - \beta DJ^{-1}\dot{Z} \tag{16}$$

for the mean motion of Z.

Now invoke Assumption 8 of Sect. 2, which I'll express explicitly as $\beta\|DJ^{-1}\| \ll 1$. Then (16) can be rewritten as

$$\dot{Z} = (I + \beta DJ^{-1})^{-1} J\nabla f \approx (J - \beta D)\nabla f. \tag{17}$$

So the result of the correction is to modify the matrix representing the Poisson bracket. D can have an antisymmetric component, which just changes J to a different antisymmetric matrix (an effect called "geometric magnetism" by [3], though it is not clear to me whether the result automatically satisfies the Jacobi identity). Assuming $\beta > 0$, the symmetric part of D, being non-negative, produces damping because under (17), $df(Z)/dt = -\beta\nabla_i f D_{ij}\nabla_j f \leq 0$ to first order.

Ruelle's formula could also be used to determine the first order change to the covariance of the fluctuations from the mean velocity, but since we are considering the fluctuations to already be first order small it does not make sense to determine this second order effect on its own.

In principle, the effects of deviations from conservation of $W_Z(E)$ should also be analysed; indeed, in the view of [25] they are responsible for the dissipation.

4.4 Effect of Autonomous Slow Motion

In reality $Z(t)$ does not move along a predetermined path but is driven by $\dot{Z} = V$, which depends on the choice of initial condition from λ_{Z_0,E_0}. The difficulties induced by back-reaction of the slow motion on the fast variables have been addressed by Kifer in some contexts [14]. It seems reasonable to add the fluctuations to the mean determined as above, but one ought to verify that correlations do not cause a further change of the same order. This produces nearly the claimed result (1), but with ∇f instead of ∇F and the interpretation of "=" unspecified.

4.5 Ergode to Monode

For large number of fast degrees of freedom, the microcanonical free energy can be approximated by the canonical free energy (up to a possible constant), assuming a thermodynamic limit exists for the heat capacity. Here is a derivation.

From (7), the canonical free energy for given β_0 satisfies $e^{-\beta_0 F(Z)} = \int e^{-\beta_0 E} W'_Z(E)\, dE$, so

$$-\beta_0 e^{-\beta_0 F}\nabla F = \int e^{-\beta_0 E}\nabla W'\, dE.$$

Integrating this by parts yields $\int \beta_0 e^{-\beta_0 E}\nabla W\, dE$, which by (6) can be written as $-\beta_0 \int e^{-\beta_0 E} W'\nabla f\, dE$, where the E-dependence of f is via $w_0 = W_Z(E)$. Thus

$$\nabla F = \int e^{-\beta_0 E} W'\nabla f\, dE \,/\, \int e^{-\beta_0 E} W'\, dE.$$

For the microcanonical ensmble at energy E, write the canonical temperature $T(E) = 1/\beta(E) = 1/(\log W')'$, from (15). For large $k = m - n$, assume the heat capacity per degree of freedom $c(\epsilon) = \frac{1}{kT'(k\epsilon)}$ as a function of the energy ϵ per degree of freedom is positive and bounded uniformly in k (say for simplicity that the limit as $k \to \infty$ exists). It follows by integration that $\beta(E) = 1/T(E)$ is a function of ϵ nearly independent of k, and by another integration the same for $s(\epsilon) = \frac{1}{k} \log W'(k\epsilon)$, the entropy per degree of freedom. Then the function $e^{-\beta_0 E} W'(E)$ of E is approximately $e^{-k(\beta_0 \epsilon - s(\epsilon))}$, which is sharply peaked around the ϵ_0 such that $s'(\epsilon_0) = \beta_0$ (because $s''(\epsilon) = -\frac{\beta^2}{c(\epsilon)} < 0$), i.e. $\beta(E_0) = \beta_0$. Thus ∇F for β_0 is approximated by ∇f for this E_0.

4.6 Klimontovich Interpretation

The Klimontovich interpretation [15] of the stochastic differential equation is as the limit as τ decreases to 0 of the discrete-time process

$$Z_i((n+1)\tau) - Z_i(n\tau) = \tau(J_{ij} - \beta D_{ij})\nabla_j F + \sigma_{ij}(Z((n+1)\tau))w_j(n) \qquad (18)$$

with $w_j(n)$ independent random steps of zero mean and variance τ. The distinguishing feature is that σ is evaluated at the end of a step rather than at the beginning (as for Ito) or averaged between the ends (as for Stratonovich). More formally, the Klimontovich interpretation is defined as the Ito equation with added drift $\nabla_j S_{ji}$ where S is the symmetric part of D.

The differences between interpretations of the stochastic differential equation are probably beyond first order, but there is a clear preference for Klimontovich interpretation, because it is the only one for which the measure $e^{-\beta F}\Omega_N$ (the marginal on N of the canonical ensemble) is stationary if S depends on Z. This can be verified by substitution in (3). It was also understood by Hänggi [6].

5 Case of Standard Mechanical System

If N is a cotangent bundle T^*L, the Hamiltonian has the form $H(Q,P,z) = \frac{1}{2}P^T M^{-1} P + h(Q,z)$ for slow degrees of freedom $(Q,P) \in T^*L$ and positive definite mass matrix $M(Q)$, and the symplectic form is $\omega = \sum_j dQ_j \wedge dP_j + \tilde{\omega}$ with $\tilde{\omega}$ a symplectic form in z (possibly depending on Q and P), then the formulae simplify. In particular $\dot{Q} = M^{-1}P$ exactly, F takes the form $F(Q,P) = \frac{1}{2}P^T M^{-1} P + G(Q)$ with G defined by $e^{-\beta G(Q)} = \int e^{-\beta h(Q,z)}\Omega$, D is independent of P, and $D_{ij} = 0$ if any index refers to a component of Q. So for the rest of this section, D denotes just its PP-block. G is called the "potential of mean force". Hence

$$d(M(Q)\dot{Q}) = -(\nabla G + \beta D\dot{Q})dt + \sigma\, dW. \qquad (19)$$

Note that this can be derived even without making the approximation used in (17).

This simpler form of Langevin equation permits a further reduction in the "over-damped" case. Specifically, if D is invertible and the maximal rate of change of M, D and G along solutions is slow compared to the slowest damping rate $\beta\|MD^{-1}\|^{-1}$ then P relaxes onto a slow manifold and is thereafter slaved to the motion of Q, resulting to a good approximation in

$$\beta D_{ij} dQ_j = -\nabla_i G(Q)\, dt + \sigma_{ij} dW_j(t), \tag{20}$$

or equivalently (with $T = 1/\beta$ and taking D symmetric for simplicity)

$$dQ = -TD^{-1}\nabla G\, dt + 2T\sigma^{-T}\, dW \tag{21}$$

(using Klimontovich interpretation again). For a proof see Theorem 10.1 of [24]. This is the form commonly used in biochemistry, e.g. [2, 21] and perhaps justifies the idea of reactions following steepest descent curves on the free energy surface. Note that it produces Smoluchowski-Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} = \mathrm{div}\left(T(D^{-1}\nabla G)\rho + T^2 D^{-1}\nabla \rho\right), \tag{22}$$

with stationary measure $e^{-\beta G} dQ$ (dQ being volume on L).

6 Quantum Degrees of Freedom

Quantum mechanics can be viewed as Hamiltonian. For Hermitian operator h on complex Hilbert space U, take the state space to be the projectivisation $P(U)$ of U (i.e. the set of 1D complex subspaces), endowed with the Fubini-Study symplectic form [22], and take $H(\psi) = \langle\psi|h\psi\rangle/\langle\psi|\psi\rangle$. The resulting Hamiltonian vector field gives Schrödinger evolution $i\frac{d\psi}{dt} = h\psi$.

Alternatively, take M to be the dual of the Lie algebra of Hermitian operators on U with inner product $\langle A,B\rangle = \mathrm{Tr}\, AB$ and its Lie-Poisson bracket, and $H(A) = \mathrm{Tr}\, hA$. This gives von Neumann evolution $idA/dt = [h,A]$.

Thus one can incorporate quantum degrees of freedom in the above framework, for example electronic degrees of freedom involved in the conformation change of rhodopsin after absorbing a photon.

There is the problem however that the quantum dynamics is not Anosov, so after all perhaps the approach of [8, 31] is more appropriate.

7 Kinetics Out of Chemical Equilibrium

The slow state space N can be a covering space. For example, its base can represent the conformation of myosin and associated momenta, and the decks differ by the number of ATP molecules. This is the appropriate way to view a biomolecular system in thermal equilibrium but out of chemical equilibrium. Indeed such problems motivated the present paper [20].

8 Conclusion and Problems

A strategy has been sketched to derive a Langevin equation for slow degrees of freedom of a Hamiltonian system, under suitable assumptions. It would be good to carry out this programme in full.

I conclude with a few related problems for the future:

- Examine the effect of the small measure exceptional set in Anosov's averaging theorem.
- Obtain higher accuracy by slightly different well selected choice of $Z(0)$.
- Determine the rank of S (degeneracies can arise only from coboundaries [4]), and conditions for hypoellipticity (e.g. [1]).
- Attempt to extend the results beyond mixing Anosov fast dynamics.
- Try to use the theory of partial hyperbolicity, because it can lead to ergodicity under fairly general circumstances [26].
- Adapt the approach for a constant pressure ensemble, e.g. by adding a heavy piston under gravity.
- Investigate what to do if there is no gap in time-scales.
- Study ways in which the result fails. This would be even more interesting than the whole current programme.

Acknowledgements This paper benefitted from the responses of participants at the "Mathematics of Model Reduction" conference in Leicester, 28–30 August 2007, and a workshop on "Microscopic origins of dissipation and noise" at the Max Planck Institute for Mathematical Sciences, Leipzig, 31 Oct–3 Nov 2007.

References

1. D. R. Bell. *Degenerate Stochastic Differential Equations and Hypoellipticity*. CRC Press, Boca Raton, FL (1995).
2. D. Bemporad, J. W. Essex, and C. Luttmann. Permeation of small molecules through a lipid bilayer: a computer simulation study. *J. Phys. Chem. B*, **108**, 4875–4884 (2004).
3. M. V. Berry and J. M. Robbins. Chaotic classical and half-classical adiabatic reactions: geometric magnetism and deterministic friction. *Proc. R. Soc. Lond. A*, **442**, 659–672 (1993).
4. J.-P. Conze and S. Le Borgne. Méthode de martingales et flot géodesique sur une surface de courbure constante n'egative. *Ergod. Th. Dyn. Sys*., **21**, 421–441 (2001).

5. D. Dolgopyat. Evolution of adiabatic invariants in stochastic averaging. *Stoch Dyn*., **4**, 265–275 (2004).
6. J. Dunkel and P. Hänggi, Theory of relativistic Brownian motion: the $(1+1)$-dimensional case. *Phys. Rev. E*, **71**, 016124 (2005).
7. G. L. Eyink, T. W. N. Haine, and D. J. Lea. Ruelle's linear response formula, ensemble adjoint schemes and Lévy flights. *Nonlinearity*, **17**, 1867–1889 (2004).
8. G. W. Ford, M. Kac, and P. Mazur. Statistical mechanics of assemblies of coupled oscillators. *J. Math. Phys*., **6**, 504–515 (1965).
9. G. Gallavotti and E. G. D. Cohen. Dynamical ensembles in stationary states. *J. Stat. Phys*., **80**, 931–970 (1995).
10. P. Gaspard. *Chaos, Scattering and Statistical Mechanics*. Cambridge University Press, Cambridge (1998).
11. D. Givon, R. Kupferman, and A. Stuart. Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity*, **17**, R55–R127 (2004).
12. T. J. Hunt and R. S. MacKay. Anosov parameter values for the triple linkage and a physical system with a uniformly chaotic attractor. *Nonlinearity*, **16**, 1499–1510 (2003).
13. C. Jarzynski. Thermalization of a Brownian particle via coupling to lowdimensional chaos. *Phys. Rev. Lett*., **74**, 2937–2940 (1995).
14. Y. Kifer. Some recent advances in averaging. In M. Brin, B. Hasselblatt, and Ya. Pesin, editors, *Modern Dynamical Systems and Applications*. Cambridge University Press, Cambridge, pp. 385–403 (2004).
15. Yu. L. Klimontovich. Nonlinear Brownian motion. *Phys. Usp*., **37**, 737–766 (1994).
16. P. Langevin. Sur la théorie du mouvement brownien. *Comptes Rendus Acad. Sci. Paris*, **146**, 530–533 (1908).
17. J. L. Lebowitz. Statistical mechanics: a selective review of two central issues. *Rev. Mod. Phys*., **71**, S346–S357 (1999).
18. E. Lieb. Some problems in statistical mechanics that I would like to see solved. *Phys. A*, **263**, 491–499 (1999).
19. P. Lochak and C. Meunier. *Multiphase Averaging for Classical Systems*. Springer, New York (1988).
20. R. S. MacKay, D. J. C. MacKay. Ergodic pumping: a mechanism to drive biomolecular conformation changes. *Phys. D*, **216**, 220–234 (2006).
21. S.-J. Marrink, H. J. C. Berendsen. Simulation of water transport through a lipid membrane. *J. Phys. Chem*., **98**, 4155–4168 (1994).
22. D. McDuff and D. Salamon. *Introduction to Symplectic Topology*. Clarendon press, Oxford (1995).
23. I. Melbourne and M. Nicol. A vector-valued almost sure invariance principle for hyperbolic dynamical systems. *Ann. Probab*., **37**, 478–505 (2009). arXiv/math.DS/0606535
24. E. Nelson. *Dynamical Theories of Brownian Motion*. Princeton University Press, Princeton (1967).
25. E. Ott. Goodness of ergodic adiabatic invariants. *Phys. Rev. Lett*., **42**, 1628–1631 (1979).
26. C. Pugh and M. Shub. Stable ergodicity. *Bull. Am. Math. Soc*., **41**, 1–41 (2004).
27. D. Ruelle. Differentiation of SRB states for hyperbolic flows. *Ergod. Th. Dyn. Sys*., **28**, 613–631 (2008). IHES/M/04/33
28. A. Soffer. Dissipation through dispersion. In I. M. Sigal and C. Sulem C, editors, *Nonlinear Dynamics and Renormalization Group*, CRM Proc Lect Notes (Am Math Soc). **27**, pp. 175–184 (2001).
29. G. I. Taylor. Diffusion by continuous movements. *Proc. Lond. Math. Soc*., **20**, 196–212 (1922).
30. M. Wilkinson. Dissipation by identical oscillators. *J. Phys. A*, **23**, 3603–3611 (1990).
31. R. Zwanzig. Nonlinear generalized Langevin equations. *J. Stat. Phys*., **9**, 215–220 (1973).

Stable Chaos

Antonio Politi and Alessandro Torcini

Abstract *Stable chaos* is a generalization of the chaotic behaviour exhibited by cellular automata to continuous-variable systems and it owes its name to an underlying irregular and yet linearly stable dynamics. In this review we discuss analogies and differences with the usual deterministic chaos and introduce several tools for its characterization. Some examples of transitions from ordered behavior to stable chaos are also analyzed to further clarify the underlying dynamical properties. Finally, two models are specifically discussed: the diatomic hard-point gas chain and a network of globally coupled neurons.

1 Introduction

Chaos is associated with an exponential sensitivity of the evolution to tiny perturbations in the initial conditions, so that the presence of at least one positive Lyapunov exponent is considered as a necessary and sufficient condition for the occurrence of irregular dynamics in deterministic dynamical systems [1]. In fact, the first observation in coupled-map models of stochastic-like behaviour accompanied by a negative maximum Lyapunov exponent came as a big surprise [2, 3]. In order to highlight the unexpected coexistence of *local stability* and *chaotic behaviour*, the phenomenon was called *stable chaos* (SC). Although the definition sounds like an oxymoron, in practice, there is no logical inconsistency, as the irregular behaviour is a transient phenomenon that is restricted to finite-time scales. In spite of this restriction, SC is both a well defined and meaningful concept, because the transient duration diverges exponentially with the system size and is therefore infinite in the thermodynamic limit. Moreover, the stationarity of SC [3] suggests that it can represent an interesting platform for studying non-equilibrium phenomena. A better understanding of

A. Politi (✉)
Istituto dei Sistemi Complessi, Consiglio Nazionale delle Ricerche;
Centro Studi Dinamiche Complesse, I-50019 Sesto Fiorentino, Italy
e-mail: antonio.politi@isc.cnr.it

M. Thiel et al. (eds.), *Nonlinear Dynamics and Chaos: Advances and Perspectives,*
Understanding Complex Systems, DOI 10.1007/978-3-642-04629-2_6,

SC can be gained by exploring the analogy with the chaotic behaviour exhibited by elementary cellular automata [4], another phenomenon that can be formally defined only in the thermodynamic limit. In fact, as we clarify in this review, SC is a sort of extension of cellular-automaton dynamics to continuous-variable systems. In particular, the spreading velocity of localized perturbations, a standard indicator used to quantify the degree of chaoticity in cellular automata, proves rather fruitful also to characterize SC. However, in this latter case it is necessary to distinguish between finite and infinitesimal perturbations (the latter ones cannot even be defined in cellular automata, because of the discreteness of the local variable) and it is thereby possible to define two conceptually different propagation velocities. This allows giving a fairly general definition of SC as that of a dynamics dominated by "finite amplitude" pertubations [5, 6].

Altogether, one can express the relevant difference between deterministic chaos and SC by referring to the relevant flux of information: while in the former context, information flows from the least towards the most significant digits, in the latter, it flows from the boundaries towards the core of the system. It would be therefore desirable to develop a general formalism able to encompass both phenomena. A promising idea is based on the introduction of "non democratic" norms which attribute increasingly small weights to the sites that are increasingly "far" from the region of interest. Although this approach allows quantifying the spatial information flow, it can be hardly extended to account for perturbations that have locally a finite amplitude, the analysis of which would require a genuine nonlinear treatment. In fact, a tool like finite amplitude Lyapunov exponents (FALEs) [7] appears to be more appropriate for characterizing SC, although it is not clear how to go beyond the maximal exponent (for the absence of a proper scalar product definition in this context).

As we have mentioned above, in systems with a finite number of degrees of freedom, SC is a transient phenomenon. One might therefore think of using tools and ideas developed for the characterization of transients such as those extensively discussed in the nice review by Tèl and Lai [8]. One must however distinguish between SC and standard chaotic transients (for a seminal paper on the subject, see [9]). In the former case, the maximum Lyapunov exponent is positive and formulas such as Kapral-Yorke and Pesin relations can be invoked to express some properties of the invariant measure in terms of the Lyapunov exponents [10]. In SC, a straightforward application of the same formulas yields manifestly useless predictions, as they do not take into account the spatial information flow that is the key mechanism of SC. Accordingly, one must still heavily rely on direct numerical simulations to infer the structure of the invariant measure. Nevertheless, we suspect that a possible common property of chaotic transients and SC is the presence of a strange repeller. In fact, chaotic transients are almost by definition the manifestation of trajectories evolving in the vicinity of a repeller, possibly characterized by a small escape rate [8, 10]. This property seems to clash with the absense of unstable orbits in most of the models exhibiting SC. However, such models are also characterized by discontinuities in phase-space and here below we argue that their smoothing gives birth to a web of unstable orbits. We are thereby lead to conjecture that even

though SC is accompanied by a negative Lyapunov exponent, its very existence requires the presence of topological chaos, i.e. of a finite topological entropy. When and whether the resulting transient dynamics is linearly stable or unstable remains however to be clarified.

The review is organized in the following way. In Sect. 2 we briefly introduce the reference models that have been mostly used to characterize SC. In Sect. 3 we properly define SC from the scaling behaviour of the transient length and discuss its properties in terms of space-time correlations and fractal dimensions. Then, in Sect. 4, we discuss the relationship with cellular automata by suitably encoding the space-time pattern. In particular, we focus our attention on the indeterminacy of the next symbol as a way to quantify the difference between the original dynamics and that of a suitable deteministic automaton rule. We also introduce and estimate the propagation velocity of localized perturbations. In Sect. 5, we compare SC with the usual deterministic chaos. This is done by smoothing an otherwise discontinuous coupled-map model and studying the dependence of standard indicators such as the maximum Lyapunov exponent on the smoothness of the dynamic rule. As a result, we identify two thresholds: (i) the first one separates the regions with postive and negative Lyapunov exponent; (ii) the second, larger, threshold separates the region where finite-amplitude perturbations propagate faster than infinitesimal ones, from that where the two velocities coincide with one another (which is the signature of a standard chaotic evolution). Moreover, we compute the multifractal spectrum of the Lyapunov exponent, showing that it has a positive tail even when the average exponent itself is negative. In Sect. 6 we discuss various order-to-chaos transitions. In fact, the analogy with cellular automata reminds us that such rules are not necessarily chaotic. The intrinsic absence of a continuous parameter makes it impossible to investigate order-to-chaos transitions in the cellular-automata context. As this restriction does not apply to SC, it makes sense and it is desirable to investigate the onset of chaotic dynamics in this latter context. We first study a coupled-map lattice, the coupling strength being the relevant control parameter. The analysis reveals the existence of a fuzzy transition region, where regular and irregular dynamics alternate in a complex manner [11]. A simple stochastic model is then introduced to gain some further insight. In the new setup, the transition is of directed-percolation type [12].

For a long time, SC has been found only in abstract mathematical models, characterized by the presence of discontinuities or nearly discontinuous[1] evolution rules. This restriction has therefore casted some doubts on the physical relevance of this phenomenon. In Sect. 7 we discuss a mechanism that can generically lead to discontinuities in physically meaningful contexts. The mechanisms requires just the presence of δ-like events such elastic collisions between particles or spike emissions by neurons. The "non-commutativity" of such events represents a genuine source of discontinuities, which may, in turn, give rise to SC. A diatomic hard-point gas and a network of coupled neurons are discussed in Sect. 7 as examples of such dynam-

[1] See the next section for a clarification of this concept.

ical systems. The neural network model allows us also to discuss an order-to-chaos transition that appears to be of "standard" type (the critical region reduces to a single point), although the universality class has not yet been identified [13]. Finally the still open problems are briefly summarised in Sect. 8.

2 Models

Most of the numerical studies of *stable chaos* have been carried out in a 1D lattice of diffusively coupled maps [14],

$$x_i(t+1) = (1-\varepsilon) f(x_i(t)) + \frac{\varepsilon}{2}\left[f(x_{i-1}(t) + f(x_{i+1}(t))\right] \tag{1}$$

where $\varepsilon \in [0{:}1]$ is the coupling constant and the map of the interval f is piecewise linear,

$$f(x) = \begin{cases} p_1 x + q_1 & 0 \le x \le x_c \\ 1 - (1-q_2)(x - x_c)/\eta & x_x < x < x_c + \eta \\ q_2 + p_2(x - x_c - \eta) & x_c + \eta < x \le 1, \end{cases} \tag{2}$$

where $x_c = (1-q_1)/p_1$. The model [5] is a continuous generalization of the systems analysed in [2, 3], which basically correspond to $\eta = 0$, i.e. to a two-branch maps. The map is continuous because the left and right limits in both connecting points do coincide $\left(f\left(x_c^-\right) = f\left(x_c^+\right), f\left(x_c^- + \eta\right) = f\left(x_c^+ + \eta\right)\right)$. Occasionally in this review we speak of "quasi-discontinuous" models, implying the presence of large but localized (in phase-space) amplifying regions. In this context, this amounts to assuming a small but non-zero width η for the middle branch. In next the section we restrict our analysis to the case $\eta = 0$.

Since this model is rather artificial (no specific physical problem lies behind the choice of f, which has been mostly selected for simplicity reasons and for coherence with the seminal paper [15]), we find it convenient to consider a second type of model, namely a chain of Duffing oscillators

$$\ddot{x}_i = -\gamma \dot{x}_i - x_i^3 + D(x_{i-1} + x_{i+1}) + (1 + G(t)\sin 2\pi t/T_1)x_i \tag{3}$$

where γ controls the dissipation, D the diffusion between nearby sites and $G(t)$ is the modulation amplitude that is periodically switched on and off; $G = A$ for mod $(t,T) < T_1$ and zero otherwise. As discussed in [4], for T_2 long enough, the Lyapunov exponent is negative, so that the evolution must eventually converge towards a periodic orbit, as it indeed does.

3 Definition and Characterization of Stable Chaos

Simulations of the above defined map have revealed the existence of long-lasting transients followed by a sudden convergence towards some periodic orbit. This suggests, and simulations confirm, that the basin of attraction of such orbits is so intricate that the convergence is exponential only for distances *homogeneously* smaller than some threshold θ. However, since there exist many different periodic orbits, one cannot estimate the transient length by determining the distance from an a priori unknown final state. One can nevertheless determine the distance $d(t,\tau)$ of the configuration $\{x_i(t)\}$ $(i = 1,N)$ at time t from any previous configuration at time $\tau < t$. As soon as there exists a τ-value such $d(t_c,\tau) < \theta$, we can conclude that $d(t+t',\tau+t')$ will tend to 0, indicating that the dynamics converges towards an orbit of period $t - \tau$. As shown in Fig. 1, the average (over different choices of the initial conditions) transient time may increase exponentially with the chain length, i.e. with the phase space dimension. This indicates that in the thermodynamic limit, the relevant dynamical regime is not the asymptotic periodic behavior that is practically unreachable, but what one would naively consider a transient regime. This scenario is reminiscent of the disordered regime in directed percolation, which, in finite systems, has necessarily a finite lifetime, as the dynamics sooner or later is absorbed by the homogeneous state [16, 17]. Irrespective of this difficulty, the disordered regime is a true "phase" in the statistical-mechanics sense, as it is stable in infinite systems, i.e. when the thermodynamic limit is taken before the infinite-time limit. It should, however, be noticed that in SC the "absorbing state" is not just a single homogenous configuration, but may be a set of different and possibly exponentially long orbits.

A second striking character of the transient is that the maximum Lyapunov exponent turns out to be negative. Like for the very existence of SC, this statement is

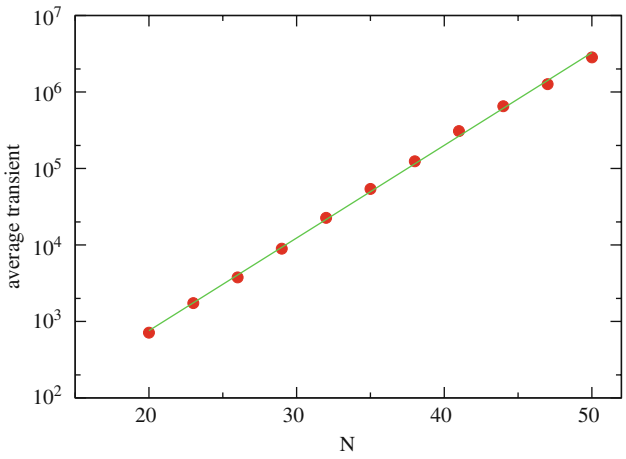


Fig. 1 Average transient duration versus the chain length for the diffusively coupled lattice of maps (2), for $p_1 = 2.7$, $q_1 = 0$, $\eta = 0$, $q_2 = 0.07$, $p_2 = 0.1$, and for the coupling strength $\varepsilon = 2/3$

formally correct only under the assumption of taking first the thermodynamic limit. In practice, it is sufficient that the transient duration is long enough to guarantee a good statistical convergence. From the data reported in Fig. 1, one can see that this is not a limitation at all, since already in a lattice of 100 maps, the periodic state is practically unreachable.

The very fact that the transient is Lyapunov-stable makes it substantially different from the chaotic transients that have been often found and attributed to the existence of some chaotic saddle of high dimensionality [8]. This is all the way more surprising once we notice that the "transient" dynamics is far from regular. In fact, simulations reveal that both spatial and temporal correlations decay exponentially. An example is reported in Fig. 2, where we plot

$$C(j) = \frac{|\langle x_i(t) x_{i+j}(t)\rangle|}{\langle x_i(t)^2\rangle} \quad , \quad C(\tau) = \frac{|\langle x_i(t) x_i(t+\tau)\rangle|}{\langle x_i(t)^2\rangle} \tag{4}$$

where $\langle\cdot\rangle$ denotes an ensemble average

It is natural to characterize the invariant measure also in terms of its fractal dimension. Since the whole Lyapunov spectrum is negative, one cannot invoke the Kaplan-Yorke [18] formula to predict the number of active degrees of freedom. Actually, such a formula would imply that the dimension is equal to zero and this is in fact true for the asymptotic attractor. Therefore, we must rely only on direct numerical computations. More precisely, we have decided to compute the correlation dimension [19] of spatial sequences of variables [20]. In other words, we have constructed embedding spaces of the type $x_i(t), x_{i+1}(t), \ldots, x_{i+_{e-1}}(t)$, for $e = 1, 2, 3$. In each case, we have counted the number of pairs of points $\mathcal{N}(e, \delta)$ that are separated by a distance larger than Δ in a space of dimension e. Afterwards, we have determined the dimension as the effective derivative, i.e.

$$D_2(e,\Delta) = \frac{\partial \log \mathcal{N}}{\partial \log \Delta} \tag{5}$$

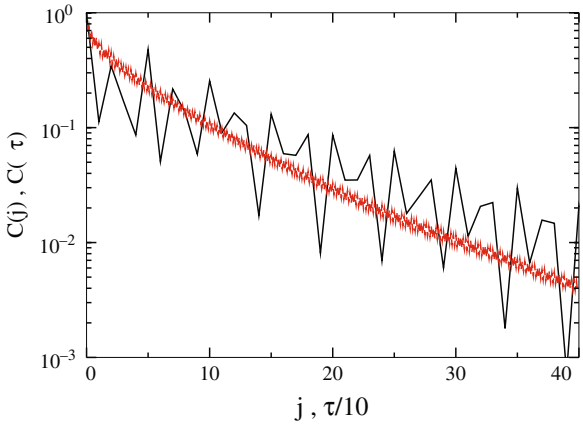


Fig. 2 Spatial and temporal (*smoother curve*) correlations for the same parameter values of the single map as in Fig. 1 and $\varepsilon = 0.608$

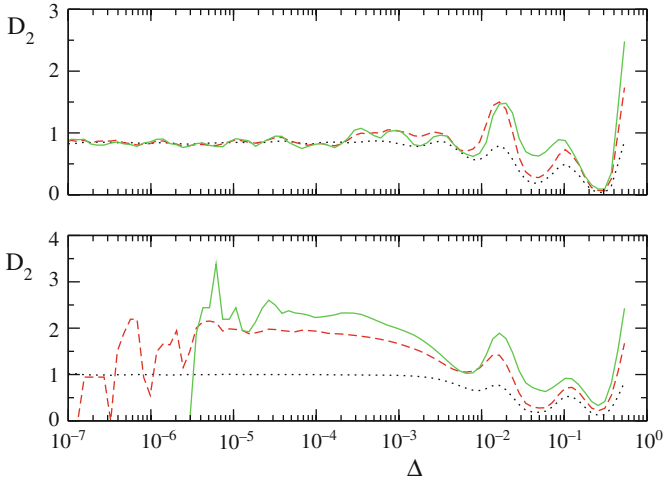


Fig. 3 Correlation dimension of the spatial embedding for $\varepsilon = 0.6008$ (*upper panel*) and $\varepsilon = 0.608$ (*lower panel*). *Dotted, dashed* and *solid curves* correspond to embedding dimension $e = 1$, 2, and 3, respectively

Formally, the correlation dimension is the limit of $D_2(e,\Delta)$ for $\Delta \to 0$. As for small Δ, $\mathcal{N}$ is affected by statistical fluctuations due to the finite number of points, the relevant question is whether the limiting behaviour sets in for distances that are numerically accessible. In Fig. 3 we report the results for two different values of the coupling strength, $\varepsilon = 0.6008$ and $\varepsilon = 0.608$, which correspond to an ordered and chaotic regime, respectively. Even in the ordered regime, the fractal dimension is finite, as revealed by the plateau, whose height is independent of the embedding dimension. The non-zero value of the dimension reflects the disordered spatial structure, i.e. the existence of *spatial chaos*. Therefore, already from this simple case we can conclude on the necessity to go beyond the standard Lypapunov-exponent analysis. In the chaotic regime, the effective dimension is larger and grows with the embedding dimension e (see lower panel in Fig. 3). However, in the absence of theoretical arguments, we cannot definitely conclude whether the dimension will saturate for $e \to \infty$, indicating the existence of a low-dimensional attractor, or whether it diverges, suggesting some form extensivity [20].

4 Relationship with Cellular Automata

The existence of a stochastic-like dynamics accompanied by an exponential contraction of infinitesimally close trajectories suggests an analogy with the so-called chaotic cellular automata (CA) rules [21]. In fact, in a finite lattice, any CA rule must eventually produce a periodic orbit, since the number of distinct states is finite, namely B^N, where B is the number of states of the local variable and N is

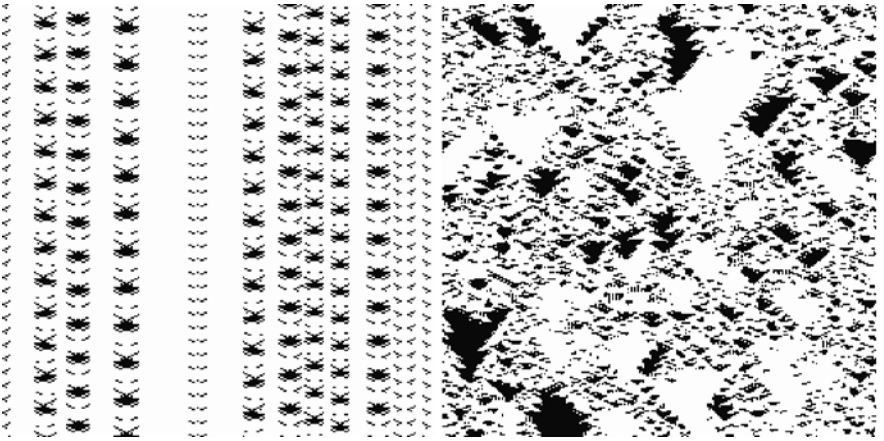


Fig. 4 Two patterns generated by iterating Eq. (1) with the function f defined as in Eq. (2) with the same parameter values as in the previous figures and coupling strength $\varepsilon = 0.55$ (*left panel*) and 0.7 (*right panel*). Time flows downwards; *black* corresponds to $x^i(t) < 1/2$

the number of lattice sites. What makes a chaotic rule different from an ordered one is precisely the time needed to cycle through previously visited states: such a time is exponentially long in chaotic rules [21].

A binary representation of the dynamics observed in the the coupled map lattice (1) confirms these expectations. The pattern plotted in Fig. 4 are indeed very reminiscent of those obtained by iterating CA rules.

The relationship with CA can be put on more firm grounds, as we discuss in the following with reference to the chain of Duffing oscillators (3). The bistable character of the single oscillators suggests a natural way to encode the underalying dynamics and thereby to explore possible connections with CA rules. An appropriate indicator to do so is the indeterminacy $\Delta h(m)$ of the symbol $s' = s_i(t+1)$, under the assumption that the sequence $S(m) = \{s_{i-r}(t), s_{i-r+1}(t), \ldots, s_i(t), \ldots, s_{i+r}(t)\}$, is observed at time t (time being measured in periods of the forcing term) and where $m = 2r + 1$. The indeterminacy is formally defined as [22]

$$\Delta h(m) = \sum_S P(S(m)) \sum_{s'} P(s'|S(m)) \log P(s'|S(m)) \tag{6}$$

where the first sum extends over all sequences of length m generated by the chain dynamics, and the second sum to the two values of the symbol s'. $P(S(m))$ is the probability to observe anywhere the sequence $S(m)$; $P(s'|S(m))$ is the conditional probability that the observation of the symbol s' at time $t+1$ on the site i is preceded by the sequence S at time t in a window of length m centered around the site i. When the knowledge of $S(m)$ allows to perfectly predict s', then the indeterminacy is zero. In this case, the symbolic dynamics is perfectly equivalent to that of a CA defined over a window of length m. In Fig. 5 we report the data for $T_2 = 8$, 18 and 20. Δh is a non decreasing function of m, since the more we assume to know on the past,

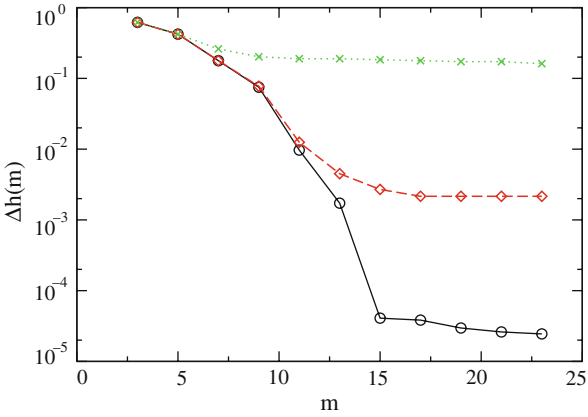


Fig. 5 Indeterminacy in the chain of Duffing oscillators (3) with $\gamma = 0.103$, $D = 0.0263$, $\omega = 0.56$, $T_1 = T_2/16$ and $T_2 = 8$, 18, 20 (from *top* to *bottom*)

the smaller must be the uncertainty on the future. If Δh becomes exactly equal to 0 for a finite m, then we can conclude that the dynamics is perfectly reconstructible from a CA with a finite interaction range. In all cases we see that for m larger than 15, the curves saturate revealing the existence of a residual uncertainty. This does neither imply that the dynamics contains some degree of stochasticity, nor that the model has to include longer memory terms. One striking such example was discussed in [22], where Crutchifield applied this approach to the pattern generated by an elementary CA, after it was suitably encoded. The indeterminacy of the encoded pattern revealed the presence of a residual uncertainty even though the CA rule is deterministic and requires only the memory of one past step, while the encoding is even memoryless. The identification of the "optimal machine" in generic cases is a typical example of the hardness of inverse problems.

The analogy with CA suggests to quantify the degree of chaos also in SC by determining the velocity v_F of propagation of perturbations. Let us consider two configurations that initially differ in the interval $[-r, r]$ and let $i_l(t)$ ($i_r(t)$) denote the leftmost (rightmost) site where they differ more than some threshold. Accordingly, we can define the front velocity as

$$v_F = \lim_{t\to\infty} \frac{i_r - i_l}{2t}. \tag{7}$$

Within CAs, a finite spreading velocity is considered as an evidence of chaotic behaviour [21]. In fact, this is true also in the context of SC, as it can be seen in Fig. 6, where we plot the spreading of an initial difference for parameter values that correspond to ordered and irregular behaviour. There we see that the perturbation spreads only in the latter case (see the right panel). In the CA language, the velocity v_f is often named the "Lyapunov exponent" [21]. In fact the evolution equation of an elementary CA can be formally written as a mapping of $\mathcal{R}^2$ into itself,

$$u^l(t+1) = F^l(u^l(t), u^r(t)) \qquad u^r(t+1) = F^r(u^l(t), u^r(t)) \tag{8}$$

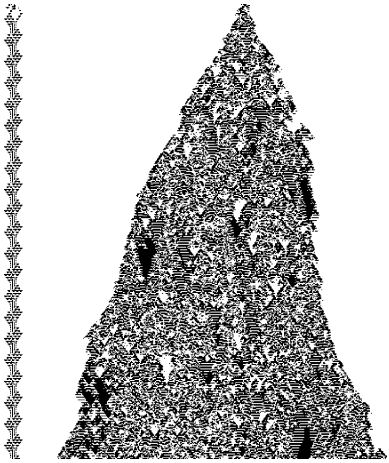


Fig. 6 Propagation of initially localized perturbations in the coupled map lattice for the same values as in Fig. 4. Time flows downwards

where $u^r(t) = \sum_{\geq 0} s_i(t)2^{-i}$, $u^l(t) = \sum_{i<0} s_i(t)2^i$ and $s_i(t) = 0,1$ (for the sake of simplicity we refer to binary automata). In general F^l and F^r are highly singular functions, but this does not forbids to define a sort of Lyapunov exponent from the growth rate of an arbitrarily small perturbation δ. In practice, if two configurations differ only in the interval $[-r, r]$, we can equivalently say that the representation points in $\mathcal{R}^2$ are separated by a distance $\delta \approx 2^{-r}$. Moreover, if the size of the spatial region where the two configurations differ increases with a velocity v_f, we can state that the $\mathcal{R}^2$ distance grows as $\exp[(v_f \log 2)t]$, thus confirming that, apart from a multiplicative factor, the velocity plays the role of a Lyapunov exponent. In coupled map lattices, the local variable is continuous rather than binary, but this does not change the substance of the argument.

5 Relationship with Deterministic Chaos

The original model where SC has been observed for the first time has a peculiarity, namely, the discontinuity of the mapping [3]. As a result, the distance between two arbirarily close trajectories can suddenly become of order $\mathcal{O}(1)$, when they find themselves on opposite sides of the discontinuity. It is therefore reasonable to study the continuous version of the model, i.e. to assume $\eta \neq 0$ in (2). In the limit $\eta \to 0$, the map (2) reduces to the original discontinuous system.

Already at the level of the single map (i.e. without invoking any spatial coupling), the introduction of an additional branch may drastically modify the structure of the corresponding dynamical system. This is clear in the simple case $q_1 = q_2 = 0$, $p_1 = p_2 = 2$. For $\eta = 0$, the topological entropy is $H = \ln 2$, as the map corresponds to the Bernoulli shift; however, for any arbitrarily small, but finite, η-value, the appearance of a third branch induces a jump to $H = \ln 3$. For the parameter values that correspond to the SC regime discussed in the previous section, the consequence

of a finite η-value is even more striking, as H is strictly equal to zero for $\eta = 0$, while it is finite for $\eta = 0^+$. This can be understood, by performing a slightly nonconventional symbolic analysis. Let us start by recalling that for the original parameter values, there exists a stable period-3 orbit (x_1, x_2, x_3), whose points are ordered as $q_2 < x_1 < x_2 < x_c < x_c + \eta < x_3 < 1$. Because of the third contracting branch, the interval [x_3,1] is asymptotically squeezed to a point, so that we can identify the rightmost point x_3, with the right border of asymptotically distinct trajectories. Analogously, the interval $[q_2, x_1]$ is also squeezed to zero and we can accordingly interpret x_1 as the left border of the relevant interval. Finally $[x_c + \eta, x_3]$ is also squeezed to zero and can be neglected as well. As a result, the relevant dynamics is described by the mapping of $I_1 = [x_1, x_2]$ and $I_2 = [x_c, x_c + \eta]$: $f(I_1) = I_2$ while $f(I_2) = I_1 \cup I_2$. It is easy to show that the corresponding topological entropy is the golden mean $H = \log[(1 + \sqrt{5})/2]$. Therefore, we can at least conclude that the introduction of a finite but arbitrarily small η induces topological chaos in an otherwise stable environment.

In the single map, the existence of a fractal chaotic repellor can induce long transients only for those trajectories that are carefully selected in the vicinity of the repellor itself. It is reasonable to conjecture that in spatially extended systems, the stable manifold of the repellor forces generic trajectories to follow an intricate arrangement of "channels" before landing on some periodic orbits. What are, however, the dynamical properties of the lattice, when finite η-values are assumed?

First of all, it important to notice that while the topological enropy jumps abruptely to a finite value, the Lyapunov exponent exhibits a smooth behaviour (see Fig. 7), i.e. for sufficiently small η-values ($\eta < \eta^* < 3.10^{-4}$) it stays negative. This implies that the phenomenon of SC is *generic* (in the mathematical sense), even though the window of existence is (at least in this context) rather narrow. More accurate information can be extracted by performing a multifractal analysis of the Lyapunov exponent [23, 24]. In particular, we have computed the probability

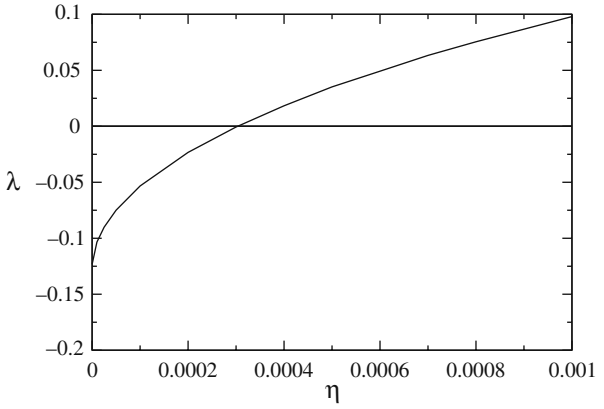


Fig. 7 Maximum Lyapunov exponent of the map lattice (2) as a function of η, while the other parameter values are the same as in Fig. 1. The results have been obtained for $N = 200$, but are practically independent of the system size

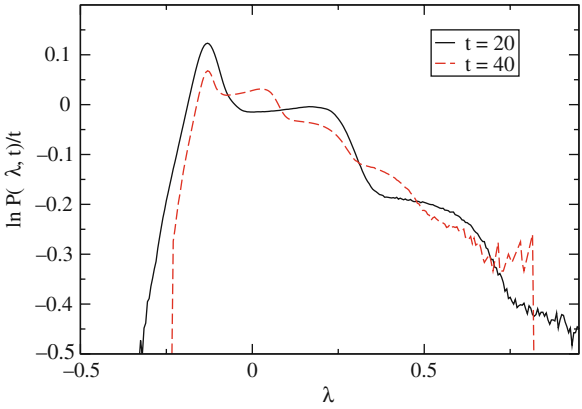


Fig. 8 Multifractal distribution of Lyapunov exponents for $\eta = 1.10^{-4}$, where the average Lyapunov exponent is still negative

distribution $P(\lambda, t)$ of the maximum Lyapunov exponent λ over a time span t. For sufficiently large t, the probability $P(\lambda, t)$ is expected to scale as $P(\lambda, t) = \exp[-G(\lambda)t]$ where G is a dynamical invariant whose operative definition is obtained by inverting this scaling relation, $G = -(\log P)/t$. In Fig. 8 we have plotted the results obtained for $t = 20$ and 40. Even though the most probable and average Lyapunov exponent is negative (the spectra refer to $\eta = 10^{-4} < \eta^*$), there is a positive tail, in agreement with the conjectured existence of a web of unstable orbits. The smoothed steps on the right of the maximum correspond to the number of times a sample trajectory is actually visiting the expanding branch. The two spectra do reasonably overlap, suggesting that the time $t = 40$ is already in the scaling regime, although finite-size corrections are still large (notice, in fat, that the maximum of G has to be, by definition, equal to 0).

Altogether, SC appears to be somehow complementary to the blow-out phenomenon discovered in the study of synchronization transitions [25]. While analysing the stability of the synchronization manifold, it has been discovered that in some circumstances, the corresponding (multifractal) Lyapunov spectrum altough mostly confined to the negative semi-axis, may exhibit a positive tail. In such a case, one has to go beyond the linear stability analysis, because whenever the distance is amplified, nonlinear terms are responsible for either bringing the trajectory back towards the manifold or letting it escape away. In the context of SC, nonlinear terms bring the trajectory back towards the "invariant manifold", although the mechanism is perfectly efficient only in the infinite dimensional limit.

In order to clarify the mechanisms by which nonlinearities contribute to stabilizing the chaotic dynamics, it is convenient to analyse the propagation of perturbations. We start by briefly recalling the concept of convective Lyapunov exponents [26]. Given a unidimensional lattice model in the stationary regime, let us introduce a δ-like perturbation at time $t = 0$ in the origin $i = 0$ and imagine to monitor the perturbation amplitude $w_i(t)$. Kaneko and Deissler [26] suggested that

$$w_i(t) = \exp[\Lambda(i/t)t] \tag{9}$$

where $\Lambda(v)$ represents the (exponential) growth rate of a perturbation in a frame moving with the velocity v. It is a priori obvious that for $v = 0$, one recovers the usual Lyapunov exponent, and that for large velocities one has to expect negative growth rates. In fact, $\Lambda(v)$ has a typically parabolic shape with the maximum in zero. All velocities for which $\Lambda(v) > 0$ correspond to growing perturbations. The limit velocity for linearly propagating perturbations is fixed by the marginal stability criterion $\Lambda(v_L) = 0$. Instead of determining directly $\Lambda(v)$, it is more convenient to exploit the chronotopic approach set in [27, 28], and formally introduce a perturbation with a spatial amplification factor

$$w_i(t) = \mathrm{e}^{-\mu i} u_i(t) \tag{10}$$

In our lattice model, the evolution rule for $u_i(t)$ reads

$$\begin{aligned} u_i(t+1,\mu) = \frac{\varepsilon}{2}\mathrm{e}^{-\mu i} f'(x_{i-1}(t)) u_{i-1}(t,\mu) + \\ (1-\varepsilon) f'(x_i(t)) u_i(t,\mu) + \frac{\varepsilon}{2}\mathrm{e}^{\mu i} f'(x_{i+1}(t)) u_{i+1}(t,\mu) \end{aligned} \tag{11}$$

By iterating this recursive equation with suitable boundary conditions (periodic conditions are typically optimal, as they reduce finite-size effects), we obtain the chronotopic growth rate $\lambda(\mu)$. Altogether, an infinitesimal perturbation $w_i(t) = \exp[\lambda(\mu)t - \mu i]$ with a spatial growth rate μ grows in time with an exponent $\lambda(\mu)$. The evolution of the initially localized perturbation is connected to $\lambda(\mu)$ by a Legendre transform

$$\Lambda(v) = \lambda(\mu) - \mu\lambda'(\mu); \quad v = \lambda' \tag{12}$$

In order to determine the velocity corresponding to a given μ-value, it is necessary to compute the derivative of $\lambda(\mu)$. Since the numerical computation of derivates is always affected by large numerical errors, it is convenient to perform a few more analytical steps [27]. By introducing,

$$u_i(t, \mu + d\mu) = u_i(t, i) + z_i(t, \mu) d\mu \tag{13}$$

in the recursive relation (11), we obtain an equation for the deviation $z_i(t, \mu)$,

$$\begin{aligned} z_i(t+1,\mu) = \frac{\varepsilon}{2}\mathrm{e}^{-\mu i} f'(x_{i-1}(t))(z_{i-1}(t,\mu) - u_{i-1}(t,\mu)) \\ + (1-\varepsilon) f'(x_i(t)) u_i(t,\mu) + \frac{\varepsilon}{2}\mathrm{e}^{\mu i} f'(x_{i+1}(t))(z_{i+1}(t,\mu) \\ - u_{i+1}(t,\mu)) . \end{aligned} \tag{14}$$

The knowledge of z_i and of u_i allows determining λ'. In fact, by taking the μ derivative in the definition of the chronotopic Lyapunov exponent,

$$\lambda(\mu) = \frac{1}{2} \lim_{t\to\infty} \frac{||\mathbf{u}(t)||^2}{t} \tag{15}$$

one obtains

$$\lambda'(\mu) = \lim_{t\to\infty} \frac{\mathbf{u}(t)\cdot\mathbf{z}(t)}{t||\mathbf{u}(t)||^2} \tag{16}$$

where $\cdot$ stands for the scalar product. In order to better understand the selection process of the propagation velocity, it is convenient to go back to the evolution of a single exponential profile $w_i(t) = \exp[\lambda(\mu)t - \mu i]$. Its velocity is obviously $V(\mu) = \lambda/\mu$. From the Legendre transform we have that

$$\frac{dV}{d\mu} = \frac{1}{\mu}\left(\frac{d\lambda}{d\mu} - \frac{\lambda}{\mu}\right) = -\frac{\Lambda}{\mu^2}\,. \tag{17}$$

Since the perturbation velocity is identified by the equation $\Lambda = 0$, we see also that it corresponds to the minimum of $V(\mu_0)$. In other words, as long as the evolution is controlled by linear mechanisms, the slowest among all possible fronts is selected. Let us now turn our attention to fronts delimiting finite perturbations. Since even such fronts must have a leading infinitesimal edge, v_F will be $v_F(\mu^*)$ for some μ^*. It is hard to imagine that μ^* is smaller than μ^0: accordingly, either $v_F = v_L$ or $v_F > v_L$. This scenario is perfectly confirmed by the study of the model (2). Solid and dashed curves in Fig. 9 correspond to v_F and V_L, respectively. There we see that v_F is strictly larger than v_L for $\eta < \eta_c \approx 1.210^{-3}$, while above η_c the two coincide within numerical accuracy. One can also notice that the linear velocity is not defined for $\eta < \eta^*$, where the system is linearly stable and no propagation of infinitesimal perturbations can occur.

As discussed in [6], the mechanism responsible for the finite difference between v_L and v_F is that perturbations of increasing amplitude (starting from infinitesimal ones) tend to propagate faster and thereby to push the corresponding front. These

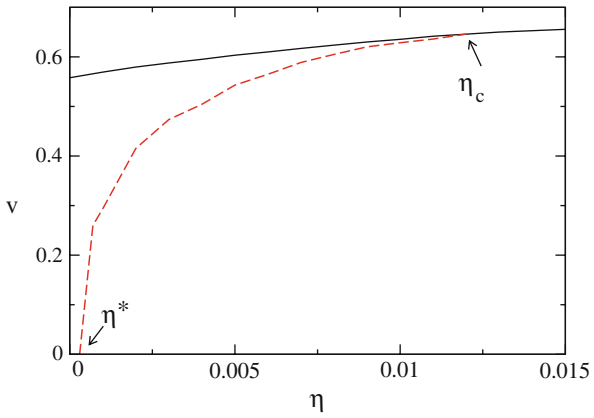


Fig. 9 Linear (v_L, *solid curve*) and front (v_F, *dashed curve*) velocity versus η. Deterministic chaos exists only for $\eta > \eta^*$. Beyond η_c, $v_L = v_F$

results are indeed fairly general and not just restricted to the model considered in this section [29]. Moreover, this phenomenology is conceptually equivalent to that observed in the context of front propagation (see, e.g. fronts connecting steady states in reaction-diffusion systems [30, 31]), that is effectively described by the famous Fisher-Kolmogorov-Petrovsky – Piskunov equation [32].

Altogether we can conclude by stating that the front velocity proves to be a useful indicator to identify the presence of SC (in spatially extended systems) from the presence of nonlinear propagation mechanisms that cannot be accounted for by linear stability analysis [6]. In such a sense, the results in Fig. 9 indicate that SC persists up to the second threshold and not just to the first one [5].

6 From Order to Chaos

Once ascertained that SC is a sort of extension of CA chaos to systems characterized by continuous variables, it is natural to investigate the possible phase transitions, a question that cannot even be posed in CAs, where all variables are discrete. The front propagation velocity v_F provides the right tool to assess the relative stability of the two phases. Let us, in fact, consider two initial conditions: a reference trajectory $\{x_i^0\}$, and a perturbed one $\{x_i\}$ differing only in a finite interval $-L < i < L$, where it is randomly set. If the interval where $v(i)$ is of order $\mathcal{O}(1)$ increases by eating the region where the field was initially equal to zero, we can conclude that the chaotic phase is thermodynamically stable.

In Fig. 10 we plot the results of careful computations performed with the coupled map lattice (1) for different values of the coupling strength ε. The ε-range has been selected so as to include both the ordered and the chaotic phase. In fact, we see that the front velocity is equal to zero (finite) in the left (right) part of the figure. However, these two clearly distinct phases are not separated by a point-like transition. We find instead a fuzzy region, where chaos and order alternate in a seemingly irregular

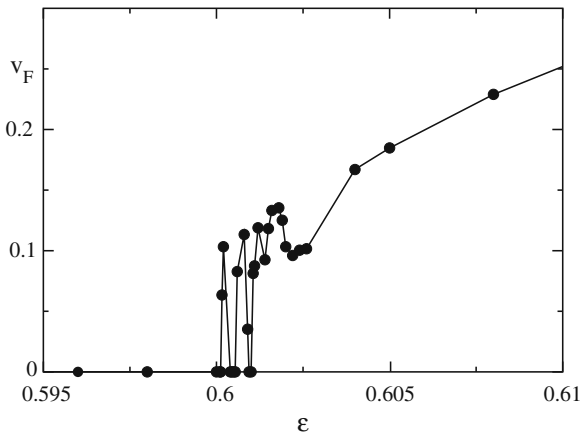


Fig. 10 Front propagation velocity in the coupled map lattice (1,2) for the same parameter values as in Fig. 1

manner. Is this an evidence of the complexity that is sometimes invoked to exist at the edge of chaos? Pure numerics alone is not sufficient to provide a convincing answer to such a difficult question.

An exact formulation and solution of this problem requires to control simultaneously two trajectories, a task that is nearly impossible. A simpler formulation which can nevertheless help to gain some insight on the transition consists in assuming a random evolution for the reference trajectory, and thus reducing the problem to that of characterizing the stochastic evolution of the difference field $v_i(t)$ [12]. In mathematical terms, this amounts to studying the equation,

$$v_i(t+1) = (1-\varepsilon)w_i(t+1) + \frac{\varepsilon}{2}[w_{i-1}(t+1) + w_{i+1}(t+1))] \tag{18}$$

where

$$w_i(t+1) = \begin{cases} v_i(t)/\eta & \text{w.p.} \quad p = a\eta \\ a v_i(t) & \text{w.p.} \quad 1-p \end{cases} \quad \text{if} \quad v_i(t) < \eta \tag{19}$$

$$w_i(t+1) = \begin{cases} 1 & \text{w.p.} \quad p = a v_i(t) \\ a v_i(t) & \text{w.p.} \quad 1-p \end{cases} \quad \text{if} \quad v_i(t) \geq \eta \tag{20}$$

The stochastic $1/\eta$ amplification simulates the effect of visiting the expanding interval of the map (2). The amplification saturates to take into account the boundedness of the dynamics. This is the only element breaking the linearity of $v_i(t)$ dynamics. Moreover, for the sake of simplicity, we assume that a uniform contraction rate a (an assumption that is basically equivalent to set $p_1 = p_2 < 1$). At variance with the original deterministic model, here a detailed numerical analysis of the parameter space (a,η), reveals that ordered and chaotic phases are separated by a standard phase transition (see Fig. 11) that belongs to the directed percolation (DP) type for

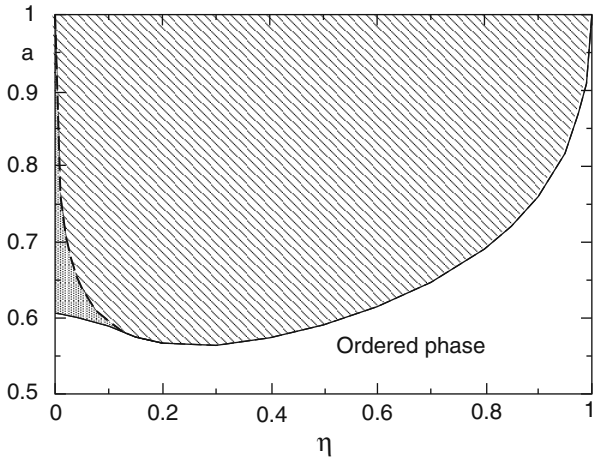


Fig. 11 Phase diagram of the stochastic model Eq. (20)

small enough values of η [33] and seems to be of multiplicative noise type beyond some critical η value [34].

Notice that this stochastic model is even more closely related to the problem of synchronization between mutually coupled map lattices (see [35–40] for a more detailed discussion), since the assumption of a stochastic evolution is appropriate everywhere in parameter space including the critical region separating the two phases.

In the SC context, the DP transition is the most relevant one, as it occurs precisely in the regime where the evolution is characterized by a negative Lyapunov exponent. DP was introduced and is usually discussed in systems where the local variable has just two states: 0, and 1. Moreover, the dynamical rule is such that 1's cannot spontaneously appear in a sea of 0's. This is the key difference with respect to the present context, where the variable v_i is continuous and thereby the 0-state is never perfectly reached (in finite times). It is therefore necessary to introduce a threshold to decide whether the 0-state has been reached, with the related problem of having to clarify whether the results are truly independent of the threshold. In order to settle this issue, we find it convenient to determine the Finite Amplitude Lyapunov Exponent (FALE) [7]. We do so by first introducing $\tau(W)$, the average time needed by the field norm

$$||w(t)|| = \frac{1}{L}\sum_{i}^{L} |w_i| \tag{21}$$

to become for the first time smaller than a preassigned threshold W.

The FALE can be thereby defined with reference to a sequence of exponentially spaced thresholds W_n ($W_n/W_{n-1} = r < 1$) as

$$\Lambda(W_n) = \frac{\log r}{\tau(W_{n+1} - W_n)}\,. \tag{22}$$

In the limit $r \to 1$

$$\Lambda(W) = \left[\frac{d\tau(W)}{d\log W}\right]^{-1}\,. \tag{23}$$

In the further limit $W \to 0$, $\Lambda(W)$ reduces to the usual Lyapunov exponent. In Fig. 12, we see that the FALE while being almost equal to zero at sufficiently large scales, becomes equal to the true Lyapunov exponent below a certain threshold W_c. Accordingly, since for $W < W_c$, the dynamics is dominated by the usual Lyapunov exponent, we can safely conclude that when the norm becomes smaller than W_c, the absorbing state will be reached with probability one and this solves the problem of an unambigious identification of the threshold. Moreover, detailed numerical simulations have revealed that W_c decreases faster than $1/L$, where L is the system size [33]. In discrete-variable systems, the minimal non-zero value

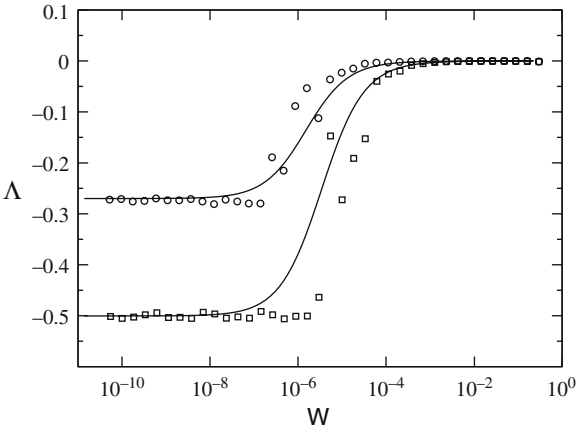


Fig. 12 Finite amplitude Lyapunov exponent of the stochastic model for two different sets of parameter values, both at criticality: *circles* refer to $\Delta = 0.01$, $L = 256$, $a_c = 0.6055$, *squares* to $\Delta = 0$, $L = 128$, and $a_c = 0.6063$

that W can meaningfully take is $1/L$ (which corresponds to just one active site). As $W_c < 1/L$, one can conclude that in this stochastic system the scaling range is even broader than in usual discrete systems. Now a comment about the reason why the linear stability analysis may not apply at vanishing distances. In fact, when the above defined norm of a vector is small, the field can nevertheless be sporadically of $\mathcal{O}(1)$. The behaviour of such bursts may represent an obstruction to the validity of the linear stability analysis and this is what tells Fig. 12.

Finally, we recall that FALE have been employed to characterize single maps of the type (2) revealing that for sufficiently small η and for some finite W, the FALE is indeed larger (positive) than the standard Lyapunov exponent [29]. Moreover, a generalization of the FALE to a comoving reference frame allows to formulate a marginal stability criterion that is able to predict the velocity on both cases of linear and nonlinear propagation [29]. Moreover, coupled maps (2) with $\eta = 0$ have been also analyzed by Letz and Kantz [41] who introduced an indicator similar to the FALE (i.e. able to quantify the growth rate of non infinitesimal perturbations). This indicator turns out to be negative for infinitesimal perturbations and becomes positive for finite perturbations. This means that a sufficiently large perturbation can propagate along the system due to nonlinear effects. This confirms previous observations for marginally stable systems [6].

7 More Realistic Models

In order to test how general *stable chaos* is, it is natural to start by asking when discontinuities or quasi-discontinuities can be expected to arise in the physical world. In fact, we have seen that the source of indeterminacy is the sudden amplification of the distance between two nearby trajectories, once they fall on opposite sides of a discontinuity. In such a case, no matter how small the initial distance is, the

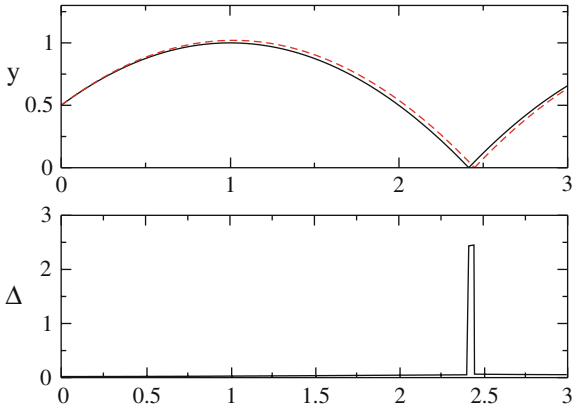


Fig. 13 Separation between two nearly equal trajectories of a point particle colliding elastically against the floor. The distance $\delta = sqrt(\delta y^2 + \delta v^2)$ is plotted versus time

separation is suddenly amplified to a value of $\mathcal{O}(1)$, that is determined by the size of the discontinuity.

Before exploring the possible occurence of such phenomena, it is important to stress that the discontinuity we are referring to is not a discontinuity in time of the type associated, e.g., to collisions. A δ-like collision induces an abrupt change of a variable (specifically, the velocity), but this affects the difference between two nearly identical trajectories only for a short (infinitesinal) time lapse, after which the trajectories come close again. This is illustrated in Fig. 13, where we have plotted the time evolution of a point-like particle bouncing elastically on the floor. In the lower panel the time evolution of the Euclidean distance is represented: only within the short time window between the collisions of the two trajectories with the floor, the relative distance becomes of order $\mathcal{O}(1)$. This is at variance with the map lattice model (1, 2), where the distance, once amplified, remains large.

In the following two subsections, we illustrate some arguments supporting the idea that a natural source of such a type of discontinuities is associated with an exchange between non-commuting δ-like events.

7.1 A Hamiltonian Model: Diatomic Hard-Point Chain

Before introducing the model, it should be remarked that in Hamiltonian systems, conservation of volumes implies that the maximum Lyapunov exponent cannot be negative; at most, all Lyapunov exponents are exactly equal to zero. In fact, the Hamiltonian version of SC is the world of marginally stable and yet ergodic models and it often goes under the name of *pseudochaos* (see, e.g. [42]). Here, we are mostly interested in emphasizing the analogies with SC and for this reason the diatomic hard point gas (HPG) turns out to be rather appropriate also for its relationship with billiard-like models, that are often invoked in the analysis of pseudochaos.

The diatomic HPG is a unidimensional system of point-like particles with masses m and M that alternate along a line and undergo elastic collisions. In the limit

$m = M$, the model is perfectly integrable, since the velocities of the two particles involved in any collision are simply interchanged. Therefore there is no mechanism leading to a diffusion in velocity space. However, as soon as the masses are assumed to be different, all numerical simulations suggest that the dynamics is ergodic. On the other hand, it is easy to convince oneself that the maximum Lyapunov exponent is still exactly equal to zero. The argument is pretty straightforward [43]: since the collision rule is linear,

$$u' = \frac{(m-M)u + 2Mv}{m+M} \tag{24}$$

$$v' = \frac{2mu - (m-M)v}{m+M} \tag{25}$$

both real- and tangent-space dynamics follow the same rule. As a result, the kinetic energy conservation ($\sum_i m_i v_i^2 = E$, where $m_i = m$ or M, depending on the parity of i) translates into the conservation of a suitably weighted Euclidean norm of the perturbation field, namely, $\sum_i m_i \delta v_i^2$. This means that the Euclidean norm of any vector is conserved, irrespective of its direction, so that all Lyapunov exponents are equal to zero.

In the absence of deterministic chaos, which is, therefore, the source of the stochastic-like behaviour exhibited by diatomic HPG chains? As illustrated in Fig. 14, we argue that the source are the discontinuities occurring around, e.g., three-body collisions. Let us consider an initial condition like that in the left panel of Fig. 14: it gives rise to a sequence of three collisions, $1-2$, $2-3$, $1-2$ before the particles separate out. By shifting the position of the central particle (this is equivalent to moving the initial x_i variable in the CML), we pass to the condition depicted in the right panel, which gives rise to the collisions $2-3$, $1-2$, $2-3$. Accordingly, the sequence of two-body collisions changes abruptly in correspondence of a three-body collision, when the three particles find themselves in the same place at the same time. As a consequence of this sudden modification, the three final velocities differ in the two cases, as it can be appreciated by comparing the two panels in

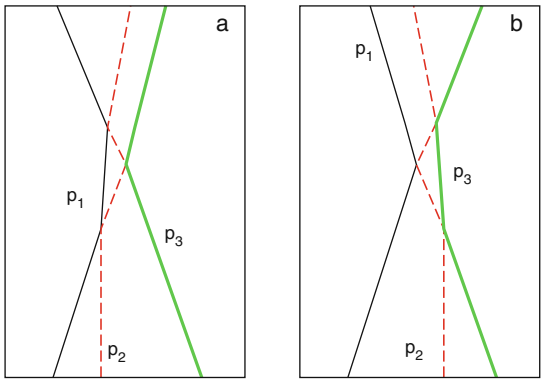


Fig. 14 Evolution of two nearly equal set of initial condition in a diatomic hard-point gas. By slightly shifting the position of the *middle* particle, the final velocities change abruptly when passing across the three-body collision

Fig. 14. Only in the limit case of equal velocities, there is no discontinuity, since the final set is the same for both sequences. In the former case, when starting from the sequence v_1, v_2, v_3, one passes first to v_2, v_1, v_3, then to v_2, v_3, v_1, and finally to v_3, v_2, v_1. One can easily verify that the final state is the same in the latter case too.

The model dynamics can be further clarified by exploring the analogy with billiard models. The connection was first discovered in [44], where the authors considered the slightly different model of hard rods. Here, we illustrate the relatively simple case of a gas of two particles P_1 and P_2 bounded to move between two fixed barriers B_l and B_r, located in $x = 0$ and $x = 1$, respectively. The linear position of the two particles can be represented as the position of a point-particle in the plane and the constraints $0 \leq x_1 \leq x_2 \leq 1$ imply that the motion is restricted to the triangular region depicted in Fig. 15. Collisions with the two mutually orthogonal edges correspond to collisions with either the left or the right barrier, while those with the diagonal correspond to interparticle collisions. Finally, the three angular points correspond to the only two possible three-body collisions, $B_lP_1P_2$, $P_1P_2B_r$ and to the synchronous occurrence of the two-body collisions B_lP_1 and P_2B_r. In the equal mass case, there is a perfect correspondence between hard point gas and the triangular billiard. Accordingly, we can invoke the conjecture raised in [45] that billiards with rational angles (expressed in π units) are necessarily ergodic. The crucial difference that appears as soon as the two masses are assumed to differ from one another is that in the billiard-like representation, the mass itself assumes a vectorial character. In particular, incoming and outgoing velocities are not mutually symmetric in correspondence of a collsion with the diagonal. However, the most relevant consequence is the appearance of true discontinuities. This is illustrated by comparing two nearby trajectories which undergo a different sequence of collisions. In the left panel of Fig. 15, which refers to equal masses, we see that the small difference in the initial velocity generates a slow linear increase of the mutual distance. In the right panel, which refers to $M = 2m$, the two trajectories, although starting from the same initial conditions, drastically separate out and find themselves very far apart after as few as 5 collisions (see the arrows).

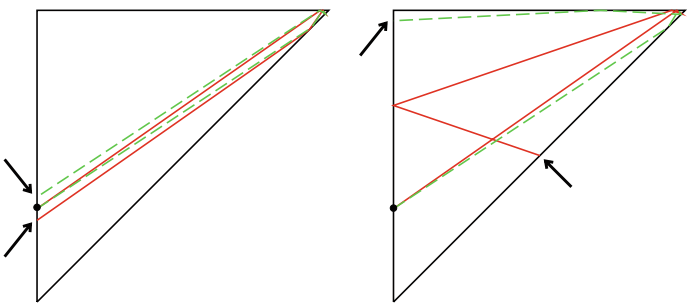


Fig. 15 Billiard-like representation of the dynamics of a chain of two particles moving in an interval with reflecting boundaries. *Horizontal* and *vertical* axes correspond to the coordinates of the first and second particle respectively. Two slightly different trajectories are plotted in each panel until the 5th collision. *Left* and *right* panels correspond to $m = M$, and $M = 2m$, respectively

7.2 Neural Networks

A perhaps more interesting example of a model exhibiting stable chaos is a network of leaky-integrate-and-fire neurons, where exponentially long transients have been identified in various set-ups [46–48, 13]. By following [13], the model dynamics for a network of N neurons can be written as a set of N differential equations

$$\dot{v}_i = c - v_i - (v_i + w) \sum_{j=1}^{N} \sum_{m} g_{ij} \delta \left(t - t_j^{(m)} \right) \tag{26}$$

where the connectivity matrix g_{ij} is defined as

$$g_{ij} = \begin{cases} G/\ell_i & \text{if } i \text{ and } j \text{ are coupled} \\ 0 & \text{otherwise,} \end{cases} \tag{27}$$

ℓ_i is the number of neurons that are connected to the ith neurons, and G is the coupling strength. Here, we will limit to consider the case of inhibitory coupling, that, in these notations, corresponds to a positive G value. All variables are dimensionless and suitably rescaled: the "membrane potential" $v_i \in [-\infty, 1]$ whenever reaches the limit value $v_j = 1$, is reset to 0 and a δ-spike is thereby emitted and received by all the connected neurons. The parameter c controls the relaxation velocity, while w quantifies indirectly the dependence of the effect of the spike on the instantaneous value of the membrane potential.

When all connections are active, the dynamics rapidly converges towards a stationary state characterized by a sequence of evenly spaced spikes (this is a so called splay state [49, 50]). In the presence of disorder, such as, e.g., a small fraction of randomly broken links, the evolution may signficantly differ, depending on the coupling strength G. Below a certain critical value, there is still a fast convergence towards an ordered state where the neurons fire in a fixed order (in agreement with Jin's theorem [51]); for sufficiently large coupling constants, the average (over different realizations of the disorder) transient length T is exponentially large with the number of neurons [13]. This is illustrated in Fig. 16, where we have plotted the average transient for different system sizes: squares and circles correspond to $G = 0.5$ and 1.8, respectively. The solid lines are the result of a linear and an exponential fit, respectively. The exponential increase of the transient is a clear indication of SC, since at the same time, the maximum Lyapunov exponent (after removing the zero exponent corresponding to a shift along the trajectory) is definitely negative (as shown in [13]).

Which is the source of such long transients? In between the spikes, the single potentials relax independently towards c (a value that is not reached, since $c > 1$). Therefore, like in the diatomic hard point gas, the evolution is piecewise linear and one can derive an analytic expression for the map as rigorously done in [13]. In the absence of jumps between different branches, the dynamics would be globally

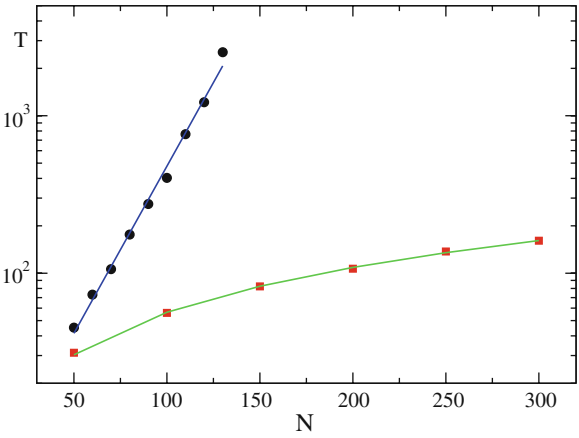


Fig. 16 Transient length for the neural network model (26) for $c = 2$, $w = 4/7$ and 5% of broken links. *Squares* (*circles*) correspond to $G = 0.5$ and ($G = 1.8$). The *solid lines* have been obtained by means of a linear and exponential fit, respectively

stable; the negativity of the Lyapunov exponent is a reminiscence of such a stability. However, like in the previous cases, there are discontinuities associated with abrupt changes in the firing order of the neurons. Let us indeed consider two neurons i and j such that $g_{ji} = 0$, while $g_{ij} \neq 0$ and consider two different initial conditions: (i) $v_i(0) = v_j(0) - \varepsilon$, (ii) $v_i(0) = v_j(0) + \varepsilon$. A schematic view of the evolution is presented in Fig. 17, where the solid line corresponds to the dynamics of the ith-neuron, while dashed and dotted line denote the former and latter trajectories, respectively. There, one can see that for times larger than t_2 the two trajectories are separated by a finite distance, as a result of a discontinuity in the dynamical law. This is due to the dependence of the inhibitory effect of a spike on the actual value of v (see the multiplicative factor $(u + w)$ in Eq. (26)). Being the size of the discontinuity of the same order of the coupling strength ($\mathcal{O}(1/N)$), one might argue that this is negligible for N large enough. This is not the case, because it has to be compared with the changes induced by the smooth dynamics in between two

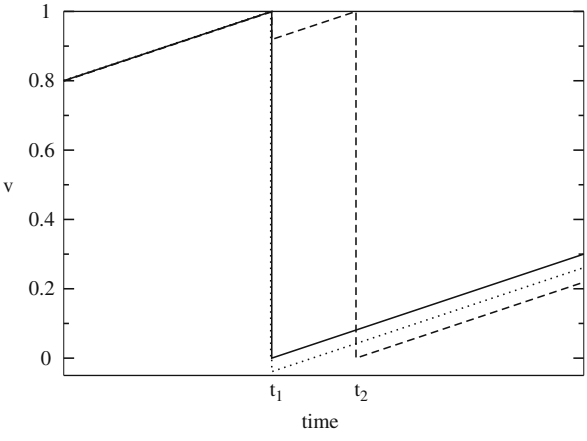


Fig. 17 Evolution of two neurons with nearly the same potential and asymmetric coupling: neuron i couples to j, but not vice versa. The *dotted line* corresponds to the evolution of neuron i; *solid* (*dashed*) *curve* corresponds to the evolution of neuron j in case its action potential is initially smaller (larger) than that of neuron j

consecutive spikes that is of the same order. Moreover, the distance between the two trajectories is even amplified to $\mathcal{O}(1)$ in the time interval $[t_1, t_2]$. As $t_2 - t_1$ is, by definition of the same order of the interspike interval, this effect too is, in principle, nonnegligible.

More recently, stable and yet irregular behaviour has been reported also in the context of a slightly different neural network, where the spike are assumed to be received with a finite delay τ and the spike effects are independent of v (see [47]). From the point of view of *discontinuities*, this latter property inhibits a persistent amplification of distances between nearby trajectories. Nevertheless, the finite-time amplification mechanism is still present and the very fact that long-stable transients have been observed is an indication that it lasts enough to yield "avalanches" and thereby to a self-sustained irregular behaviour. However, one should also notice that "discontinuities" are a necessary but not sufficient condition for the onset of SC.

8 Conclusions

In the present Review we have thoroughly discussed the phenomenon of *stable chaos*, a type of irregular behaviour occurring in deterministic systems that manifests itself as an exponentially (with the system size) long and stationary transient. SC differs from usual chaos in that it is characterized by negative Lyapunov exponents, but still reatins some features that are reminiscent of deterministic chaos. In fact, by smoothing out the discontinuities present in the most typical SC models, induces the multifractal spectrum of the maximum Lyapunov exponent to extend to positive values. This, in turn, suggests that topological chaos (i.e. a strictly positive topological entropy) is a prerequisite for the observation of SC. However, we have shown that linear stability analysis does not to provide a convincing description of relevant properties such as the propagation of finite perturbations. In this respect, a promising indicator is represented by the finite amplitude Lyapunov exponent, although there are conceptual difficulties in extending this approach beyond the maximum exponent.

A further interesting question concerns the transition from ordered behaviour to SC. A detailed numerical analysis of a coupled-map lattice reveals the existence of a fuzzy region, where behaviour that is neither strictly ordered nor clearly chaotic has been detected. Is this just a difficulty due to strong finite size effects, or this phenomenon hides the presence of a genuinely "complex" (uncomputable) evolution? In an almost globally-coupled neural network, the transition appears to be a standard point-like phenomenon, whose universality class is however still unclear.

The most important question concerns the generality of SC. All models where SC has been observed do possess strong localized nonlinearities that may reduce to true discontinuities in phase space. The first models where SC has been observed are somehow artificial systems with no direct relationship with the physical world. However, the discussion of the diatomic hard point gas and of the network of pulse coupled neurons, has contributed to clarify that discontinuities may spontaneously

emerge in models characterized by the presence of non communting "δ-like" events (such as two-body collisions or spike emissions). Moreover, since we have seen that SC survives a smoothing of the coupled-map model, we may also conjecture that the same holds true in these latter contexts, once we assume finite collision times or finite pulse-widths.

Acknowledgements We would like to acknowledge those who have collaborated with us on this problem over the years: R. Bonaccini, F. Cecconi, M. Cencini, F. Ginelli, P. Grassberger, R. Kapral, S. Lepri, R. Livi, G.L. Oppo, and R. Zillmer. Moreover, we wish to thank M. Timme, M. Wolfrum, and S. Yanchuk for recent useful discussions. This work has been partly carried out with the support of the EU project NEST-PATH-043309.

References

1. E. N. Lorenz. Deterministic nonperiodic flow. *J. Atmos. Sci*., **20**, 130 (1963).
2. J. P. Crutchfield and K. Kaneko. Are attractors relevant to turbulence ? *Phys. Rev. Lett*., **60**, 2715 (1988).
3. A. Politi, R. Livi, G. L. Oppo, and R. Kapral. Unpredictable behaviour in stable systems. *Europhys. Lett*., **22**, 571 (1993).
4. R. Bonaccini and A. Politi. Chaotic-like behaviour in chains of stable nonlinear oscillators. *Physica D*, **103**, 362 (1997).
5. A. Politi and A. Torcini. Linear and nonlinear mechanisms of information propagation. *Europhys. Lett*., **28**, 545 (1994).
6. A. Torcini, P. Grassberger, and A. Politi. Error propagation in extended systems. *J. Phys*., **A27**, 4533 (1995).
7. E. Aurell, G. Boffetta, A. Crisanti, G. Paladin, and A. Vulpiani. Growth of noninfinitesimal perturbations in turbulence. *Phys. Rev. Lett*., **77**, 1262 (1996); Predictability in the large: an extension of the concept of Lyapunov exponent. *J. Phys*., **A30**, 1 (1997).
8. T. Tél and Y. -C. Lai. Chaotic transients in spatially extended systems. *Phys. Rep*., **460**, 245 (2008).
9. C. Grebogi, E. Ott, and J. A. Yorke. Fractal basin boundaries, long-lived chaotic transients, and unstable-unstable pair bifurcation. *Phys. Rev. Lett*., **50**, 935 (1983).
10. E. Ott. *Chaos in Dynamical Systems*. Cambridge University Press, Cambridge (1993).
11. F. Cecconi, R. Livi, and A. Politi. Fuzzy phase transition in a 1D coupled stable-map lattice. *Phys. Rev. E*, **57**, 2703 (1998).
12. F. Ginelli, R. Livi, and A. Politi. Emergence of chaotic behaviour in linearly stable systems. *J. Phys. A Math. Gen*., **35**, 499 (2002).
13. R. Zillmer, R. Livi, A. Politi, and A. Torcini. Desynchronization in diluted neural networks. *Phys. Rev. E*, **74**, 036203 (2006).
14. I. Waller and R. Kapral. Spatial and temporal structure in systems of coupled nonlinear oscillators. *Phys. Rev. A*, **30**, 2047 (1984); K. Kaneko. Period-doubling of Kink–Antikink patterns, quasiperiodicity in antiferro-like structures and spatial intermittency in coupled logistic lattice. *Prog. Theor. Phys*., **72**, 980 (1984).
15. L. A. Bunimovich, R. Livi, G. Martinez-Mekler, and S. Ruffo. Coupled trivial maps. *Chaos*, **2** 283 (1992).
16. P. Grassberger. Directed percolation: Results and open problems. In: S. Puri and S. Dattagupta (Eds.) *Nonlinearities in Complex Systems*. Narosa Publishing House, New Delhi (1997).
17. H. Hinrichsen. Nonequilibrium critical phenomena and phase transitions into absorbing states. *Adv. Phys*., **49**, 815 (2000).

18. J. L. Kaplan and J. A. Yorke. Chaotic behavior of multidimensional difference equations. *Lect. Not. Math*., **13**, 730 (1979).
19. P. Grassberger and I. Procaccia. Characterization of strange attractors. *Phys. Rev. Lett*., **50**, 346–349 (1983).
20. P. Grassberger. Information content and predictability of lumped and distributed dynamical systems. *Phys. Scr*., **40**, 346 (1989).
21. S. Wolfram. *Theory and Applications of Cellular Automata*, Advanced Series on Complex Systems. World Scientific, Singapore (1986).
22. J. P. Crutchfield. Unreconstructible at any radius. *Phys. Lett. A*, **171**, 52 (1992).
23. R. Benzi, G. Paladin, G. Parisi, and A. Vulpiani. On the multifractal nature of fully developed turbulence and chaotic systems. *J. Phys. A Math. Gen*., **17**, 3521 (1984).
24. G. Boffetta, M. Cencini, M. Falcioni, and A. Vulpiani. Predictability: a way to characterize complexity. *Phys. Rep*., **356**, 367 (2002).
25. A. Pikovsky, M. Rosenblum, and J. Kurths. *Synchronization : A Universal Concept in Nonlinear Sciences*. Cambridge University Press, Cambridge (2002).
26. R. J. Deissler and K. Kaneko. Velocity-dependent Lyapunov exponents as a measure of chaos for open-flow systems. *Phys. Lett*., **A119**, 397 (1987).
27. A. Politi and A. Torcini. Periodic orbits in coupled Henon maps: Lyapunov and multifractal analysis. *Chaos*, **2**, 293 (1992).
28. S. Lepri, A. Politi, and A. Torcini. Chronotopic Lyapunov analysis: (I) a comprehensive characterization of 1D systems. *J. Stat. Phys*., **82**, 1429 (1996); Chronotopic Lyapunov analysis: (II) towards a unified approach. *J. Stat. Phys*., **88**, 31 (1997); Entropy potential and Lyapunov exponents. *Chaos*, **7**, 701 (1997).
29. M. Cencini and A. Torcini. Linear and nonlinear information flow in spatially extended systems. *Phys. Rev. E*, **63**, 056201 (2001).
30. W. van Saarloos. Front propagation into unstable states: Marginal stability as a dynamical mechanism for velocity selection. *Phys. Rev. A*, **37**, 211 (1988); Front propagation into unstable states. II. Linear versus nonlinear marginal stability and rate of convergence. *Phys. Rev. A*, **39**, 6367 (1989).
31. U. Ebert and W. van Saarloos. Front propagation into unstable states: universal algebraic convergence towards uniformly translating pulled fronts. *Phys. D*, **146**, 1 (2000).
32. R. A. Fisher. The wave of advance of advantageous genes. Ann. Eugenics, **7**, 355 (1937); A. N. Kolmogorov, I. Petrovsky, and N. Piscounov. Etude de l'equation de la diffusion avec croissance de la quantité de matiére et son application a un probleme biologique. *Bull. Univ. Moscow Ser. Int*., **A1**, 1 (1937).
33. F. Ginelli, R. Livi, A. Politi, and A. Torcini. On the relationship between directed percolation and the synchronization transition in spatially extended systems. *Phys. Rev. E*, **67**, 046217 (2003).
34. M. A. Muñoz. Multiplicative noise in non-equilibrium phase transitions: a tutorial. In: E. Korutcheva and R. Cuerno (Eds.) *Advances in Condensed Matter and Statistical Mechanics*. Nova Science Publishers, New York, p. 37 (2004).
35. L. Baroni, R. Livi, and A. Torcini. Transition to stochastic synchronization in spatially extended systems. *Phys. Rev. E*, **63**, 036226 (2001).
36. V. Ahlers and A. Pikovsky. Critical properties of the synchronization transition in space-time chaos. *Phys. Rev. Lett*., **88**, 254101 (2002).
37. F. Bagnoli and F. Cecconi. Synchronization of non-chaotic dynamical systems. *Phys. Lett. A*, **282**, 9 (2001).
38. F. Ginelli, V. Ahlers, R. Livi, D. Mukamel, A. S. Pikovsky, A. Politi, and A. Torcini. From multiplicative noise to directed percolation in wetting transitions. *Phys. Rev. E*, **68**, 065102(R) (2003).
39. F. Bagnoli and R. Rechtman. Synchronization universality classes and stability of smooth coupled map lattices. *Phys. Rev. E*, **73**, 026202 (2006).
40. M. Cencini, C. J. Tessone, and A. Torcini. Chaotic synchronizations of spatially extended systems as non-equilibrium phase transitions. *Chaos*, **18**, 037125 (2008).

41. T. Letz and H. Kantz. Characterization of sensitivity to finite perturbations. *Phys. Rev. E*, **61**, 2533 (2000).
42. P. Castiglione, M. Falcioni, A. Lesne, and A. Vulpiani. *Chaos and Coarse Graining in Statistical Mechanics*. Cambridge University Press, Cambridge (2008).
43. P. Grassberger, W. Nadler, and L. Yang. Heat conduction and entropy production in a one-dimensional hard-particle gas. *Phys. Rev. Lett*., **89**, 180601 (2002); P. Grassberger, W. Nadler, and L. Yang. Heat conduction and entropy production in a one-dimensional hard-particle gas. nlin.CD/0203019v1
44. S. L. Glashow and L. Mittag. Three rods on a ring and the triangular billiard. *J. Stat. Phys*., **87**, 937 (1997).
45. G. Casati and T. Prosen. Mixing property of triangular billiards. *Phys. Rev. Lett*., **83**, 4729 (1999).
46. A. Zumdieck, M. Timme, T. Geisel, and F. Wolf. Long chaotic transients in complex networks. *Phys. Rev. Lett*., **93**, 244103 (2004).
47. S. Jahnke, R. -M. Memmesheimer, and M. Timme. Stable irregular dynamics in complex neural networks. *Phys. Rev. Lett*., **100**, 048102 (2008).
48. R. -M. Memmesheimer. *Precise spike timing in complex neural networks*. PhD thesis, Gottingen University, Göttingen (2007).
49. S. H. Strogatz and R. E. Mirollo. Splay states in globally coupled Josephson arrays: analytical prediction of Floquet multipliers. *Phys. Rev. E*, **47**, 220 (1993).
50. R. Zillmer, R. Livi, A. Politi, and A. Torcini. Stability of the splay state in pulse-coupled networks. *Phys. Rev. E*., **76**, 046102 (2007).
51. D. Z. Jin. Fast convergence of spike sequences to periodic patterns in recurrent networks. *Phys. Rev. Lett*., **89**, 208102 (2002).

Superpersistent Chaotic Transients

Ying-Cheng Lai

Abstract Superpersistent chaotic transients are characterized by the following scaling law for its average lifetime: $\tau \sim \exp[C(p-p_c)^{-\chi}]$, where $C > 0$ and $\chi > 0$ are constants, $p \geq p_c$ is a bifurcation parameter, and p_c is its critical value. As p approaches p_c from above, the exponent in the exponential dependence diverges, leading to an extremely long transient lifetime. Historically the possibility of such transient raised the question of whether asymptotic attractors are relevant to turbulence.

Superpersistent chaotic transients were first discovered by Grebogi et al. in 1983. In their seminal work, unstable – unstable pair bifurcation was identified as the dynamical mechanism for the transients. In this Review this bifurcation and how it leads to superpersistent chaotic transients will be described. The occurrence of the transients in spatially extended dynamical systems will then be exemplified. Superpersistent chaotic transients associated with the riddling bifurcation that creates a riddled basin of attraction will be discussed, and the effect of noise on the transient lifetimes will be addressed. Finally, application to a physical problem, advection of finite-size particles in open hydrodynamical flows, will be demonstrated.

1 Introduction

Chaotic transients in low-dimensional dynamical systems are typically characterized by an algebraic scaling law of its average lifetime τ with some parameter variation [1, 2]:

$$\tau \sim (p-p_c)^{-h}, p > p_c, \tag{1}$$

Y.C. Lai (✉)
School of Electrical, Computer and Energy Engineering, Department of Physics, Arizona State University, Tempe, Arizona 85287, USA; Institute for Complex Systems and Mathematical Biology, School of Natural and Computing Sciences, King's College, University of Aberdeen, UK
e-mail: Ying-Cheng.Lai@asu.edu

M. Thiel et al. (eds.), *Nonlinear Dynamics and Chaos: Advances and Perspectives,* Understanding Complex Systems, DOI 10.1007/978-3-642-04629-2_7,

where $h > 0$ is the algebraic scaling exponent. There exists, however, another distinct class of transient chaos – superpersistent chaotic transients that are characterized by the following scaling law for their average lifetime [3, 4]:

$$\tau \sim \exp\left[C(\Delta p)^{-\chi}\right], \tag{2}$$

where $\Delta p = p - p_c$, p is a system parameter, $C > 0$ and $\chi > 0$ are constants. As p approaches the critical value p_c from above, the transient lifetime τ becomes superpersistent in the sense that the exponent in the exponential dependence diverges. This type of chaotic transients was conceived to occur through the dynamical mechanism of unstable – unstable pair bifurcation, in which an unstable periodic orbit in a chaotic attractor collides with another unstable periodic orbit on the basin boundary [3, 4]. The same mechanism causes the riddling bifurcation [5] that creates a riddled basin [6], so superpersistent chaotic transients can be expected at the onset of riddling. The transients were also identified in a class of coupled-map lattices, leading to the speculation that asymptotic attractors may not be relevant for turbulence [7]. Noise-induced superpersistent chaotic transients were demonstrated [8] in phase synchronization [9] of weakly coupled chaotic oscillators. Signatures of noise-induced superpersistent chaotic transients were also found [10] in the advective dynamics of inertial particles in open fluid flows [11].

Section 2 describes unstable – unstable pair bifurcation and explains why the bifurcation can lead to a superpersistent chaotic transient. Section 3 demonstrates the presence of the chaotic transient at the riddling bifurcation. The next topic is superpersistent chaotic transient in a coupled-map lattice system (Sect. 4). The phenomenon of noise-induced superpersistent chaotic transients is described in Sect. 5. An application to advective dynamics of inertial particles in open chaotic flows is presented in Sect. 6.

2 Unstable – Unstable Pair Bifurcation

Unstable – unstable pair bifurcation has been identified as the generic mechanism for superpersistent chaotic transients [3–5]. One can imagine two unstable periodic orbits of the same periods, one on the chaotic attractor and another on the basin boundary, as shown in Fig. 1a. In a noiseless situation, as the bifurcation parameter p reaches a critical value p_c, the two orbits *coalesce* and disappear simultaneously, leaving behind a "channel" in the phase space through which trajectories on the chaotic attractor can escape, as shown in Fig. 1b. The chaotic attractor is thus converted into a chaotic transient, but the channel created by this mechanism is typically extremely narrow [3–5]. Suppose on average, it takes time T for a trajectory to travel through the channel in the phase space so that it is no longer on the attractor, we expect T to be infinite for $p = p_c$ but, for $p > p_c$ the time becomes finite and decreases as p is increased from p_c. For p above but close to p_c, the tunneling time can be long. As we will argue below, we expect T to increase at least algebraically as Δp is decreased.

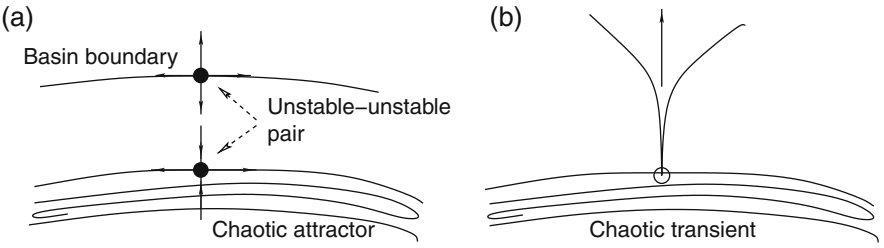


Fig. 1 (**a**) For $p < p_c$, a chaotic attractor, the basin boundary, and the pair of unstable periodic orbits. (**b**) For $p > p_c$, escaping channel created by an unstable-unstable pair bifurcation that converts the originally attracting motion into a chaotic transient

From Fig. 1a, we see that if the phase space is two dimensional, the periodic orbit on the attractor is a saddle and the one on the basin boundary is a repeller. This can arise if the map is noninvertible. Thus, the unstable-unstable pair bifurcation can occur in noninvertible maps of at least dimension two, or in invertible maps of at least dimension three (or in flows of dimension of at least four).

Let $\lambda > 0$ be the largest Lyapunov exponent of the chaotic attractor. After an unstable – unstable pair bifurcation the opened channel is locally transverse to the attractor. In order for a trajectory to escape, it needs to spend at least time $T(\Delta p)$ at the location of the opening on the attractor centered about the *mediating* periodic orbit involved in the bifurcation, stipulating that the trajectory must come to within distance of about $\exp[-\lambda T(\Delta p)]$ from this orbit. The probability for this to occur is proportional to $\exp[-\lambda T(\Delta p)]$. The average time for the trajectory to remain on the attractor, or the average transient lifetime, is thus

$$\tau \sim \exp[\lambda T(\Delta p)]. \tag{3}$$

We see that the dependence of $T(\Delta p)$ on Δp, which is the average time that trajectories spend in the escaping channel, or the *tunneling time*, is the key quantity determining the scaling of the average chaotic transient lifetime τ.

To obtain the scaling dependence of the tunneling time $T(\Delta p)$ on Δp, we note that, since the escaping channel is extremely narrow, for typical situations where $\lambda > 0$ and $T(\Delta p)$ large, the dynamics in the channel is approximately one dimensional along which the periodic orbit on the attractor is stable but the orbit on the basin boundary is unstable for $p < p_c$ (Fig. 1a). This feature can thus be captured through the following simple one-dimensional map:

$$x_{n+1} = x_n^2 + x_n + p, \tag{4}$$

where x denotes the dynamical variable in the channel and p is a normalized bifurcation parameter with critical point $p_c = 0$ (we thus write $T(p)$). For $p < p_c = 0$, the map has a stable fixed point $x_s = -\sqrt{-p}$ and an unstable fixed point $x_u = \sqrt{-p}$.

These two collide at p_c and disappear for $p > p_c$, mimicking an unstable-unstable pair bifurcation.

Since $T(p)$ is large, map (4) can be approximated in continuous-time as

$$\frac{dx}{dt} \approx x^2 + p. \tag{5}$$

Suppose the root of the channel is at $x = 0$ and its length is l. The tunneling time is given by

$$T(p) \approx \int_0^l \frac{dx}{x^2 + p} \sim p^{-1/2}. \tag{6}$$

Substituting Eq. (6) into Eq. (3), we obtain

$$\tau(p) \sim \exp(C_0 p^{-1/2}), \tag{7}$$

where $C_0 > 0$ is a constant. We see that as p approaches the critical value $p_c = 0$ from above, the average transient lifetime diverges in an exponential-algebraic way, giving rise to a superpersistent chaotic transient.

To give a concrete example, we use the class of two-dimensional maps constructed by Grebogi et al. [3, 4]:

$$\begin{aligned} \theta_{n+1} &= 2\theta_n \bmod 2\pi, \\ z_{n+1} &= az_n + z_n^2 + \beta \cos\theta_n, \end{aligned} \tag{8}$$

where a and β are parameters. Because of the Z_n^2 term in the z-equation, for large z_n we have $|z_{n+1}| > |z_n|$. There is thus an attractor at $z = +\infty$. Near $z = 0$, depending on the choice of the parameters, there can be either a chaotic attractor or none. For instance, for $0 < \beta \ll 1$, there is a chaotic attractor near $z = 0$ for $a < a_c = 1 - 2\sqrt{\beta}$ and the attractor becomes a chaotic transient for $a > a_c$ [3]. The chaotic attractor, its basin of attraction, and part of the basin of the infinity attractor are shown in Fig. 2.

Following the same argument leading to the scaling law (7), one can see that the map system (8) allows for superpersistent for $a > a_c$. In particular, for $a < a_c$ there are two fixed points: $(\theta_1, z_1) = (0, z_b)$ and $(\theta_2, z_2) = (0, z_c)$, where $z_{c,b} = (1 - a \pm r)/2$ and $r = \sqrt{(1-a)^2 - 4\beta}$. The fixed points $(0, z_b)$ and $(0, z_c)$ are on the basin boundary and on the chaotic attractor, respectively. They coalesce at $a = a_c$. For $a > a_c$, a channel is created through which trajectories on the original attractor can escape to the attractor at infinity. At the location of the channel where $\theta = 0$, the z-mapping can be written as

$$z_{n+1} - z_n = (a - 1)z_n + z_n^2 + \beta.$$

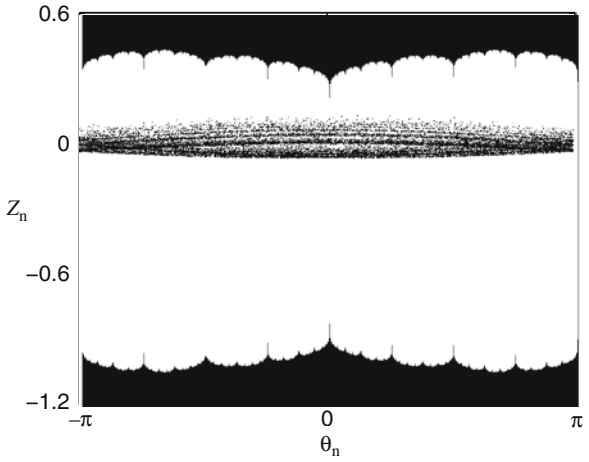


Fig. 2 For the two-dimensional map model (8), a chaotic attractor near $z = 0$, its basin of attraction (*blank*), and the basin of the attraction of the attractor at $z = +\infty$ (*black*). From [3] with permission

Letting $\delta = z - z_*$, where z_* is the minimum of the quadratic function of z on the right-hand side, we have

$$\delta_{n+1} = \delta_n + \delta_n^2 + b, \tag{9}$$

where $b \equiv \sqrt{\beta}(a - a_c) + [(a - a_c)/2]^2$. For $a > a_c$, we have $b \approx \sqrt{\beta}(a - a_c)$. In the continuous-time approximation, the dynamics in δ can be described by $d\delta/dt = \delta^2 + b$. Thus the time T required to tunnel through the escaping channel is

$$T \approx \frac{1}{b^{1/2}} \int_0^\infty \frac{d\delta}{\delta^2 + 1} = \frac{\pi}{2b^{1/2}}.$$

Since the θ-dynamics is uniformally chaotic with Lyapunov exponent $\lambda = \ln 2 > 0$, the probability for a trajectory to fall in the opening of the channel and to stay near there in the θ-direction for consecutively T iterations is proportional to $e^{-T \ln 2}$. For $a > a_c$, the average chaotic transient time is thus given by

$$\tau \sim e^{T \ln 2} \approx e^{(\pi \ln 2/2) b^{-1/2}} \approx \exp\,[C(a - a_c)^{-1/2}], \tag{10}$$

where $C = \pi(\ln 2)\beta^{-1/4}/2$ is a positive constant. Thus for $a > a_c$ the algebraic exponent in the scaling of the tunneling time with the parameter variation assumes the value of $1/2$, as shown in Fig. 3.

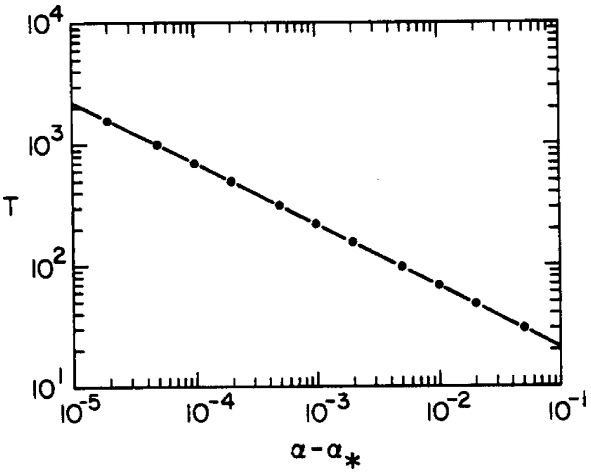


Fig. 3 For the two-dimensional map model (8), an example of the numerically obtained scaling law for superpersistent chaotic transient. From [3] with permission

3 Riddling Bifurcation and Superpersistent Chaotic Transients

The presence of symmetry in a dynamical system often leads to an invariant subspace where, in the absence of symmetry-breaking or random perturbations, a trajectory originated in the invariant subspace remains there forever. Situations can also arise where a chaotic attractor lies in the invariant subspace. One common example is the system of coupled, identical chaotic oscillators. The synchronization manifold is naturally a low-dimensional invariant subspace in the full phase space. If another attractor exists outside the invariant subspace, riddling can occur in the sense that the basin of the chaotic attractor in the invariant subspace is riddled with holes of all sizes that belong to the basin of the other attractor. Imagine the situation where all unstable perioic orbits embedded in the chaotic attractor are stable with respect to perturbations in the direction transverse to the invaraint subspace. In this case, almost all initial conditions in the vicinity of the invariant subspace lead to trajectories that end up asymptotically on the chaotic attractor. Riddling bifurcation [5] refers to the situation where, when a system parameter changes, an unstable periodic orbit (usually of low period) [12] embedded in the chaotic attractor, becomes transversely unstable. As pointed out in Ref. [5], an immediate physical consequence of the riddling bifurcation is that, when there is a small amount of symmetry-breaking, an extraordinarily low fraction of the trajectories in the invariant subspace diverge. This means that a typical trajectory would spend an extremely long time in the vicinity of the chaotic attractor before approaching the other coexisting attractor. The average lifetime of the chaotic transient versus the amount of symmetry-breaking was shown [5] to obey the scaling law for superpersistent chaotic transients.

In a two-dimensional phase space, the invariant subspace is a line. In this case, the onset of riddling is determined by a saddle-repeller bifurcation (eigenvalue +1) [3, 4]. A chaotic attractor in the invariant line is typically one-dimensional. Let $\mathbf{x}_p$ be

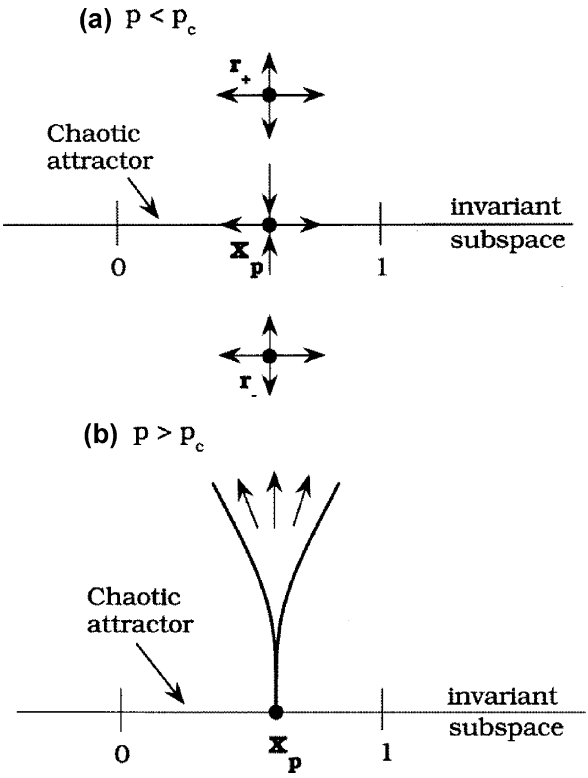


Fig. 4 (**a**) Unstable saddle fixed point in the invariant subspace and two repellers off the invariant subspace for $p < p_c$ (before the saddle-repeller pitchfork bifurcation). (**b**) Tongue structure formed for $p > p_c$, after the onset of riddling. Trajectories originated from initial conditions inside the tongues escape the invariant subspace

an unstable fixed point embedded in the chaotic attractor in the invariant subspace. The unstable point is stable transversely to this subspace, as shown in Fig. 4a. Riddling occurs when $\mathbf{x}_p$ loses its transverse stability as a parameter p passes through a critical value p_c. The loss of transverse stability is induced by the collision at $p = p_c$ of two repellers $\mathbf{r}_+$ and $\mathbf{r}_-$, located symmetrically with respect to the invariant subspace, with the saddle at $\mathbf{x}_p$ (a saddle-repeller pitchfork bifurcation). These two repellers exist only for $p \leq p_c$, as shown in Fig. 4a. For $p > p_c$, the saddle $\mathbf{x}_p$ becomes a repeller, and the two original repellers $\mathbf{r}_+$ and $\mathbf{r}_-$ off the invariant subspace no longer exist.

Due to nonlinearity, a "tongue" opens at $\mathbf{x}_p$, allowing trajectories near the invariant subspace to escape for $p > p_c$, as shown in Fig. 4b. Each preimage of $\mathbf{x}_p$ also developes a tongue simultaneously. Since the preimages of $\mathbf{x}_p$ are dense in the invariant subspace, an infinite number of tongues open simultaneously at $p = p_c$, indicating that initial conditions arbitrarily close to the invariant subspace can approach another attractor. Trajectories in the chaotic attractor, however, remain there even for $p > p_c$, since the subspace in which the chaotic attractor lies is invariant and each tongue has a zero width there. But trajectories near the chaotic attractor have a finite probability of being in the open and dense set of tongues. The basin of attraction for the chaotic attractor is then a Cantor-set of leaves of positive Lebesgue measure,

signifying riddling. Physically, since the onset of riddling induces the supernarrow tongues near the invariant subspace, superpersistent chaotic transient arises [3, 4].

To give an example, consider the following general class of dynamical systems [5]:

$$\begin{aligned} \mathbf{x}_{n+1} &= \mathbf{f}(\mathbf{x}_n), \\ \mathbf{y}_{n+1} &= \sigma + pg(\mathbf{x}_n)\mathbf{y}_n + \text{high order odd terms of } \mathbf{y}_n, \end{aligned} \tag{11}$$

where $\mathbf{x} \in R^N$ ($N \geq 1$), $\mathbf{y} \in R^M$ ($M \geq 1$), $\mathbf{f}(\mathbf{x}_n)$ is a map with a chaotic attractor in the invariant subspace $\mathbf{y}_n = 0$, $g(\mathbf{x}_n) = 1$ at some unstable periodic orbit of $\mathbf{f}(\mathbf{x}_n)$, p is a system parameter, and $pg(\mathbf{x}_n)$ is assumed to be positive. We call $\sigma \geq 0$ the symmetry-breaking parameter. Notice that for initial $y_0 \geq 0$ and $\sigma \geq 0$, trajectories have $y_n \geq 0$ for all times. The following two-dimensional version of Eq. (11) can be analyzed [5]:

$$\begin{aligned} x_{n+1} &= ax_n(1 - x_n), \\ y_{n+1} &= \sigma + pe^{-b(x-x_p)^2} y_n + y_n^3, \end{aligned} \tag{12}$$

where, for $\sigma = 0$, $y = 0$ defines the invariant subspace as a trajectory with $y = 0$ will remain so forever, and a,$b > 0$ are parameters. The symmetry-breaking parameter is with respect to the symmetry $y \to -y$. Thus, the dynamics in the invariant subspace is described by the logistic map $x_{n+1} = ax_n(1 - x_n)$ for which chaotic attractors occur for parameter values in a positive Lebesgue measure set [13].

An analysis similar to these in Sect. 2 leads to the following scaling law for the average transient lifetime with respect to the symmetry-breaking parameter σ:

$$\tau \sim \exp\left[K\sigma^{-2/3}\right], \tag{13}$$

where K is a positive constant proportional to the Lyapunov exponent of the chaotic logistic map. An example of the scaling is shown Fig. 5, a plot of $\log_{10}\tau$ versus $\sigma^{-2/3}$. Note that the exponent $2/3$ is a consequence of the y^3 term in the $y-$dynamics. If the y^3 term is replaced by, say, a y^2 term, then the exponent would be $1/2$. Thus, the exponent $2/3$ is specific to the two-dimensional map model Eq. (12).

The escaping behavior of trajectories, once they have fallen into the tongue, can be seen by monitoring their traces in the phase space before they reach $y = 1$. Since the tongues are supernarrow at $p = p_c$, it is numerically convenient to examine the case where $p > p_c$. Figure 6 shows the last 50 points for 600 trajectories before they reach $y = 1$, where $p = 1.18$ and $\sigma = 0.005$. It can be seen there is a "mushroom-shape" (tongue) crowd of trajectory points in the phase space located above the fixed point x_p. The thick solid curves in Fig. 6 represent the envelope of the tongue. These curves can be derived analytically by considering the escaping dynamics in the vicinity of x_p. Specifically, after a trajectory falls into the escaping channel located at x_p, its dynamics can be approximated by: (1) $(x_{n+1} - x_p) \approx (2 - a)(x_n - x_p)$;

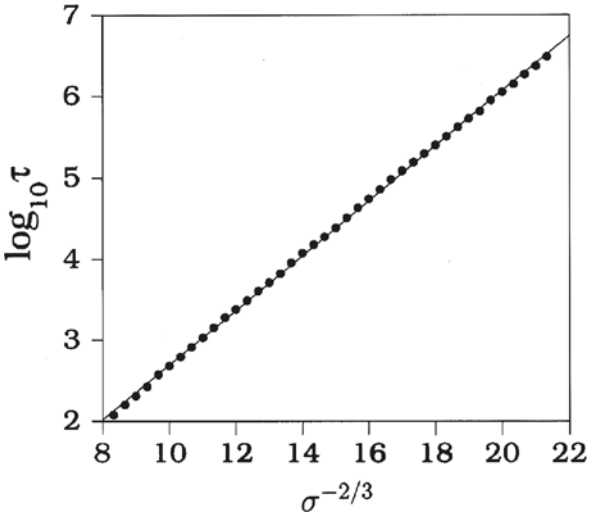


Fig. 5 Average transient time τ versus the symmetry-breaking parameter σ for $0.01 \leq \sigma \leq 0.04$ at $p = p_c = 1$. The parameters are $a = 3.8$ and $b = 5.0$ in Eq. (12). The plot is $\log_{10} \tau$ versus $\sigma^{-2/3}$. From [5] with permission

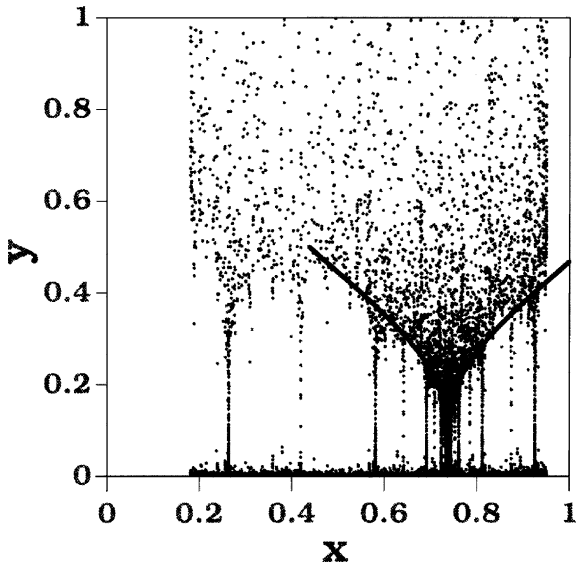


Fig. 6 *Mushroom-shaped* phase-space regions (tongues) through which trajectories escape the $y = 0$ chaotic attractor for $p = 1.18$ and $\sigma = 0.005$. From [5] with permission

and (2) $y_{n+1} \approx \sigma + py_n + y^3$. Let $z_n \equiv |x_n - x_p|$, the $x-$dynamics becomes: $z_{n+1} = |2 - a|z_n = (a - 2)z_n$. For p close to p_c and σ small, the $z-$ and $y-$ dynamics can be approximated by

$$\begin{aligned} dz/dt &= \rho z, \\ dy/dt &= \sigma + (p-1)y + y^3, \end{aligned} \tag{14}$$

where $\rho = a - 3$. The following formula for the edge of the tongue for $\sigma = 0$ and $p > 1$ can then be obtained:

$$z = \left(\frac{y}{\sqrt{(p-1) + y^2}} \right)^{\rho/(p-1)}, \text{ for } p > p_c = 1. \tag{15}$$

The solid curves in Fig. 6 are $(x_p \pm z)$, respectively. In principle, additional terms such as cxy (c is constant) in the dy/dt equation should be considered, but this has negligible effect on the results [5]. In more general cases where the system does not have a skew-product structure, one should also consider terms in the $x-$equation such as σy, σy^2 or even higher-order terms in y. But for y small (near the invariant subspace) one has $\sigma y \ll y$. Thus, these terms have negligible effect on the properties of the escaping tongues.

4 Superpersistent Chaotic Transients in Spatiotemporal Systems

An approach to studying spatially extended dynamical system is to examine various spatial patterns and their dynamical evolution. In a turbulent state, the pattern evolution appears random but statistical quantities usually converge for all practical time scales [14]. Situations can occur where, after a long time, the system falls onto a low-dimensional attractor. In this case, the high-dimensional, turbulent behavior may be only a transient. It is not possible to determine whether the observed "turbulence" is transient unless the asymototic time regime is reached. If the transient time is much longer than any physically realizable time, the system is effectively turbulent, regardless of the nature of the asymptotic attractor. In this sense, attractors are not relevant to turbulence. Crutchfield and Kaneko [7] recognized the possibility of extremely long transient in spatiotemporal dynamical systems. They demonstrated, by using a prototype model, that the attractor can typically be low-dimensional but the transient dynamics can be high-dimensional and complicated. As the system size N is increased, the transient time can grow exponentially or even faster, as follows:

$$\tau_N \sim \exp(CN^{\alpha}), \tag{16}$$

where $C > 0$ and $\alpha \geq 1$. We see that the transient is superpersistent in the limit $N \to \infty$. The case of $\alpha \leq 1$ where the growth of the transient time is exponential or slower with N was referred to as *type-I transient turbulence*, while the case $\alpha > 1$ as *type-II transient turbulence* [7]. One example of Type-I transient turbulence is chaotic defect motion in coupled map lattices where the relaxation time for disappearance of the complex patterns increases at most exponentially with the system size. For type-II transient turbulence, the pattern evolution typically appears turbulent and high-dimensional.

To demonstrate superpersistent chaotic transients in spatially extended dynamical systems, Crutchfield and Kaneko [7] used the following coupled-map-lattice

model in which both time and space are discrete but the dynamical variables are continuous:

$$x_{n+1}^{i} = \frac{1}{2r+1} \sum_{j=-r}^{r} f\left(x_n^{i+j}\right), \; i = 0, \ldots, N-1, \tag{17}$$

where n and i are discrete time and space, respectively, $f(x)$ is a nonlinear map governing the local dynamics, and r is a parameter defining the range of spatial coupling. For nearest-neighbor coupling, $r = 1$. Crutchfield and Kaneko chose the following piecewise linear map, the "dripping handrail" model, for $f(x)$:

$$f(x) = sx + \omega \; (\mathrm{mod} 1), \tag{18}$$

where s and ω are parameters. The local dynamics thus consists of an increase of ω with each iteration but when the dynamical variable x exceeds a threshold $x_{drop} = (1-\omega)/s$, a sudden decrease from unity occurs. Physically, the coupled map lattice system Eq. (17) represents a simplified model of a dripping fluid layer, where the local map $f(x)$ models the dynamics of an isolated drop. The map $f(x)$ can actually generate complicated dynamics such as chaos and it was also used to study the dynamics of the stirred Belousov – Zhabotinsky chemical reaction [15].

Crutchfield and Kaneko suggested that both type-I and type-II transient turbulence are due to the complex, hierarchical phase space structure and the transient relaxation can be regarded as a sequence of transitions through a hierarchy of subbasins. These subbasins are subspaces of a basin separated by walls through which a trajectory cannot pass except at portals. In particular, for type-I transient turbulence, the phase space is organized as a hierarchy of subbasins of decreasing dimension. Patterns near the attractor move in relatively low-dimensional subbasins, while those far away from the attractor in high-dimensional subbasins. The collision and annihilation of two defects correspond to an orbit moving from one subbasin to another. Some constant spatial length L can then be defined for the portal, which is determined by the defect size and the local geometry of the annihilation process. The phase-space volume of the portal is thus $V \sim c^L$, where $c < 1$ is the relative size of the portal with respect to the size of the subbasin. Since the number of defects in a random initial pattern is proportional to N, the probability for the sequence of transitions down through the hierarchy is $P_N \sim c^{NL}$. Assuming the dynamics within each subbasin is ergodic, the average transient lifetime is

$$\tau_N \sim P_N^{-1} \sim c^{-NL} \equiv (\bar{c})^{NL}, \tag{19}$$

where $\bar{c} = 1/c > 1$. For type-II transient turbulence, numerical evidence suggests that the patterns are generally complex during the transient epoch but occasionally they can be quite uniform. This implies that the underlying subbasins may consist of long tenrils passing through the neighborhood of the final attractor that corresponds to a simple, uniform pattern. That is, a trajectory can be relatively close to the attractor at some time but most times it moves away from it in order to find the correct path to actually reach the attractor. The subbasin hierarchy can be approximated by a direct product of the local basin structure at each spatial site. The number

of subbasins is proportional to N^γ, where γ measures the density of the tendrils. Since there are no localized annihilation events, passage through a portal is spatially global. The probability p of passing a portal is thus $p \sim c^N$. The total probability of passing all portals to reach the final attractor is the product of N^γ such local probabilities. The average transient time is

$$\tau_N \sim c^{-N^{1+\gamma}} \equiv (\bar{c})^{N^{1+\gamma}}, \tag{20}$$

which increases faster than exponentially with system size.

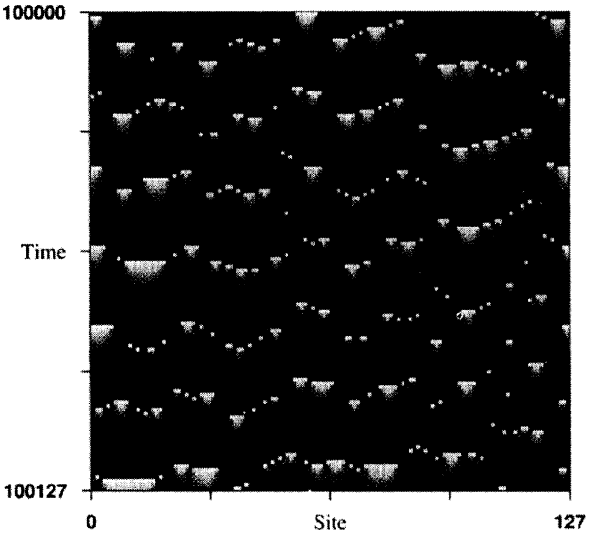


Fig. 7 For the coupled-map lattice system Eqs. (17) and (18), a typical space-time diagram with site amplitude $\{x_n^i\}$ plotted from *black* to *white* for a 128-site, spatially periodic lattice. From [7] with permission

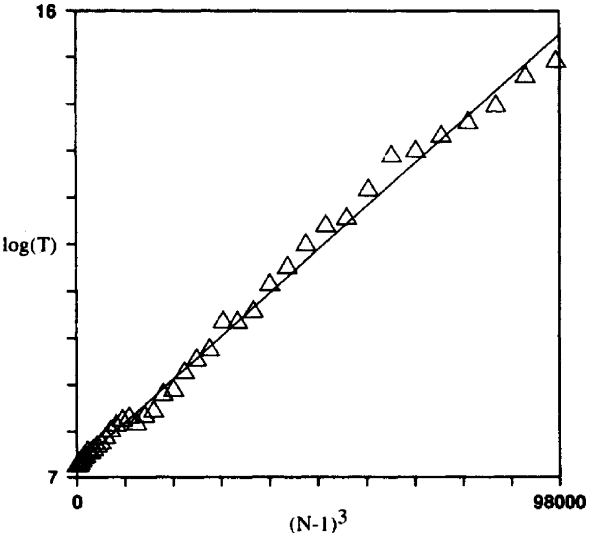


Fig. 8 Example of hyperexponential scaling of the average transient lifetime for the coupled-map lattice system Eqs. (17) and (18). From [7] with permission

A typical evolution of the lattice from a random initial condition is shown in Fig. 7, which is somewhat irregular but it eventually settles down to a uniform pattern. The scaling of the average transient lifetime in the coupled-map lattice system Eqs. (17) and (18) is hyperexponential, as shown in Fig. 8, where $\log \tau_N$ versus $(N-1)^3$ is plotted (together with a linear fitting) for $N = 1, \ldots, 47$. The hyperexponential growth in the transient lifetime means that transient spatial chaos is in principle unsimulatable on finite-state machines for lattices of large size. For turbulence this means that the final attractor may never be observed and thus irrelavant for practical purposes.

5 Noise-Induced Superpersistent Chaotic Transients

In the general setting where an unstable-unstable pair bifurcation occurs, noise can induce superpersistent chaotic transients *preceding* the bifurcation. Consider, in the noiseless case, a chaotic attractor and its basin of attraction. When noise is present, there can be a nonzero probability that two periodic orbits, one belonging to the attractor and another to the basin boundary, get close and coalesce temporally, giving rise to a nonzero probability that a trajectory on the chaotic attractor crosses the basin boundary and moves to the basin of another attractor. Transient chaos thus arises. Due to noise, the channels through which trajectory escapes the chaotic attractor open and close intermittently in time. The probability of escape is extremely small because escaping through the channel requires staying of the trajectory in a small vicinity of the opening of the channel consecutively for a finite amount of time, which is an event with extremely small probability. In this sense, the channel must be "super" narrow [3–5], leading to a superpersistent chaotic transient. The creation of the channel by noise and the noisy dynamics in the channel are thus the key to understanding the noise-induced transient behavior.

There are two regimes of interest. In the subcritical case, there is a chaotic attractor and no escaping channel exists in the absence of noise. In this case, the channel is induced by noise and it opens and closes randomly in time. In the supercritical case, the channel is open and there is already a superpersistent chaotic transient. The presence of noise affects the deterministic dynamics in the channel. In both cases, the dynamics in the channel can be regarded as being driven by a stochastic force and, hence, it can be modeled by a stochastic differential equation, the solution to which gives the tunneling time through the channel. Apparently, this time depends on the noise amplitude. The dependence, in combination with the small probability for a trajectory to move to the opening of the channel and to stay there for the duration of the tunneling time, gives the scaling of the average lifetime of the superpersistent chaotic transients with the noise amplitude.

Let ε be the noise amplitude. To obtain the scaling dependence of the tunneling time $T(\varepsilon)$ on ε, we use the following one-dimensional map:

$$x_{n+1} = x_n^{k-1} + x_n + p + \varepsilon\xi(n), \tag{21}$$

where $k \geq 3$ is an odd integer so as to generate a pair of fixed points with different unstable dimension, $p_c = 0$, and $\xi(n)$ is a Gaussian random process of zero mean and unit variance. If the tunneling time is $T \gg 1$, Eq. (21) can be approximated by

$$\frac{dx}{dt} = x^{k-1} + p + \varepsilon\xi(t), \tag{22}$$

For $p < 0$, the deterministic system for Eq. (22) has a stable fixed point $x_s = -|p|^{1/(k-1)}$ and an unstable fixed point $x_u = |p|^{1/(k-1)}$, but there are no more fixed points for $p > 0$, as shown in Fig. 9. Let $x_r = x_s$ for $p < 0$ and $x_r = 0$ for $p \geq 0$, and let T_p^k be the tunneling time. A properly formulated first-passage-time problem for this one-dimensional stochastic process yields the scaling of T_p^k with the noise amplitude ε [16, 17].

Let $P(x,t)$ be a probability density function of the stochastic process governed by Eq. (22), which satisfies the Fokker – Planck equation [18, 19]:

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial}{\partial x}[(x^{k-1} + p)P(x,t)] + \frac{\varepsilon^2}{2}\frac{\partial^2 P}{\partial x^2}. \tag{23}$$

Let l be the effective length of the channel in the sense that a trajectory with $x > l$ is considered to have escaped the channel. The time required for a trajectory to travel through the channel is equivalent to the mean first passage time T from x_r to l. Focusing on a trajectory that escapes eventually, we assume that, once it falls into the channel through x_r, it will eventually exit the channel at $x = l$ without returning

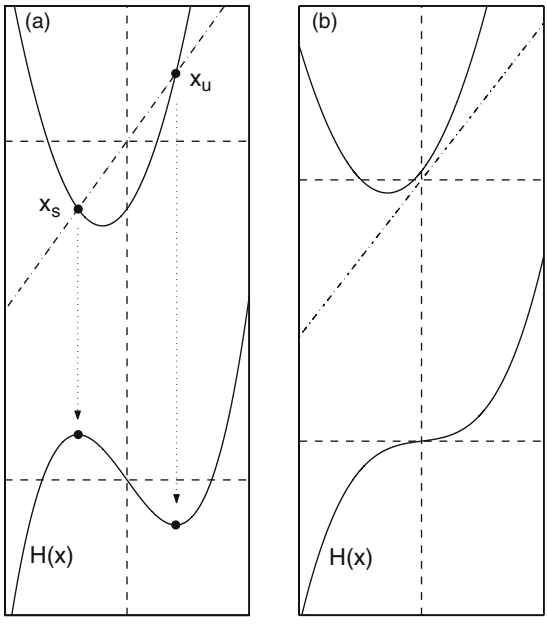


Fig. 9 For the prototype model Eq.+ (21), (**a**) the stable and unstable fixed points for the subcritical cases ($p < 0$), and (**b**) the supercritical case ($p > 0$). From [17] with permission

to the original chaotic attractor. This is reasonable considering that the probability for a trajectory to fall in the channel and then to escape is already exponentially small and, hence, the probability for any "second-order" process to occur, where a trajectory falls in the channel, moves back to the original attractor, and falls back in the channel again, is negligible. For trajectories in the channel there is thus a reflecting boundary condition at $x = 0$:

$$\left[P(x,t) - \frac{\partial P}{\partial x}\right]|_{x=x_r} = 0. \tag{24}$$

That trajectories exit the channel at $x = l$ indicates an absorbing boundary condition at $x = l$:

$$P(l,t) = 0. \tag{25}$$

Assuming that trajectories initially are near the opening of the channel (but in the channel), we have the initial condition

$$P(x,x_r) = \delta\left(x - x_r^+\right). \tag{26}$$

Under these boundary and initial conditions, the solution to the Fokker – Planck equation yields the following mean first-passage-time [18, 19] for the stochastic process (22):

$$T_p^k(\varepsilon) = \frac{2}{\varepsilon^2}\int_{x_r}^{l} dy \exp\left[-bH(y)\right]\int_{x_r}^{y} \exp\left[bH(y')\right]dy' \tag{27}$$

where $H(x) = (x^k + kpx)$ and $b = 2/(k\varepsilon^2)$.

The double integral in Eq. (22) can be carried out [16, 17] for the three distinct cases: critical ($p = 0$), supercritical ($p > 0$), and subcritical ($p < 0$). The results can be summarized as follows.

- For the small noise regime ($\varepsilon \ll \varepsilon_c \sim |p|^{k/(2(k-1))}$),

$$T_p^k(\varepsilon) \sim \begin{cases} p^{-(k-2)/(k-1)}, & p > 0, \\ \varepsilon^{-(2-4/k)}, & p = 0, \\ |p|^{-(k-2)/(k-1)} \exp\left(\frac{|p|^{k/(k-1)}}{\varepsilon^2}\right), & p < 0. \end{cases} \tag{28}$$

- For the large noise regime ($\varepsilon \gg \varepsilon_c$),

$$T_p^k(\varepsilon) \sim \varepsilon^{-(2-4/k)}. \tag{29}$$

These laws imply the following scaling laws for the average lifetime of the chaotic transients in various regimes:

- For the small noise regime ($\varepsilon \ll \varepsilon_c \sim |p|^{k/(2(k-1))}$),

$$\tau_p^k(\varepsilon) \sim \begin{cases} \exp\left[p^{-(k-2)/(k-1)}\right], & p > 0, \\ \exp\left[\varepsilon^{-(2-4/k)}\right], & p = 0, \\ \exp\left(|p|^{-(k-2)/(k-1)} \exp\left[|p|^{k/(k-1)}/\varepsilon^2\right]\right), & p < 0. \end{cases} \tag{30}$$

- For the large noise regime ($\varepsilon \gg \varepsilon_c$), we have

$$\tau_p^k(\varepsilon) \sim \exp\left(\varepsilon^{-(2-4/k)}\right). \tag{31}$$

The general observation is that for large noise ($\varepsilon \gg \varepsilon_c$), the transient is normally superpersistent. For small noise, three behaviors arise depending on the bifurcation parameter p: constant (independent of noise) for the supercritical regime, normally superpersistent for the critical case, and extraordinarily superpersistent for the subcritical regime in the sense of scaling in (30) (for $p < 0$). Numerical support for these distinct scaling behaviors can be found in [16, 17].

6 Application: Advection of Inertial Particles in Open Chaotic Flows

The phenomenon of superpersistent chaotic transients finds application in fluid dynamics. It has been known that ideal particles of zero mass and size follow the velocity of the flow and, as such, the advective dynamics can be described as Hamiltonian [20, 21] in the physical space for which chaos can arise but not attractors. In an open Hamiltonian flow, ideal particles coming from the upper stream must necessarily go out of the region of interest in finite time. However, the inertia of the advective particles can alter the flow locally [22]. As a result, the underlying dynamical system becomes dissipative for which attractors can arise and, hence, particles can be trapped permanently in some region in the physical space [23, 24]. This phenomenon was demonstrated in a model of two-dimensional flow past a cylindrical obstacle [11]. This result has implications in environmental science where forecasting aerosol and pollutant transport is a basic task, or even in defense where the spill of a toxin or biological pathogen in large-scale flows is of critical concern. The possibility that toxin particles can be trapped in physical space is particularly worrisome. It is thus interesting to study the structural stability of such attractors [10]. In particular, can chaotic attractors so formed be persistent under small noise? It was found [10] that in general, the attractor is destroyed by small noise and replaced by a chaotic transient, which is typically superpersistent. For small noise, the extraordinarily long trapping time makes the transient particle motion practically equivalent to an attracting motion with similar physical or biological effects.

This finding suggests a way to directly observe superpersistent chaotic transients in laboratory experiments.

For an ideal, passive particle of zero inertia and zero size advected in a flow, the particle velocity **v** is the flow velocity **u** which, in a two-dimensional physical space, is determined by a stream function $\Psi(x,y,t)$: $u_x = \partial\Psi/\partial y$ and $u_y = -\partial\Psi/\partial x$. For particles of finite size, viscous friction arises and, as such, their velocities differ from those of the fluid. Consider a spherical particle of radius a and mass m_p, and fluid of dynamic viscosity μ and element mass m_f, the equation of motion of the advective particle is [22]

$$m_p \frac{d\mathbf{v}}{dt} = m_f \frac{d\mathbf{u}}{dt} - (m_f/2)\left(\frac{d\mathbf{v}}{dt} - \frac{d\mathbf{u}}{dt}\right) - 6\pi a\mu(\mathbf{v} - \mathbf{u}), \tag{32}$$

where on the right-hand side, the first term is the fluid force from the undisturbed flow field, the second term is the force due to the added mass effect, and the third represents the Stokes drag. While in principle, the fluid velocity **u** is disturbed by the particle motion, if the particle sizes are relatively small and their concentration is low, **u** can be considered as unchanged [11]. For convenience, one can introduce the mass ratio parameter

$$R = \frac{2\rho_f}{\rho_f + 2\rho_p}$$

and the inertial parameter

$$A = \frac{R}{\frac{2}{9}(a/L)^2 R_e},$$

where ρ_p and ρ_f are the densities of the particle and of the fluid, respectively, L is a typical large-scale mixing length, and R_e is the Reynolds number. The equation of motion can then be casted into a dimensionless form. To simulate random forcing due to flow disturbance or other environmental factors, we add terms $\varepsilon\xi_x(t)$ and $\varepsilon\xi_y(t)$ to the force components in the x- and y-directions, where $\xi_x(t)$ and $\xi_y(t)$ are independent Gaussian random variables of zero mean and unit variance, and ε is the noise amplitude. The final equation of motion under random perturbations is

$$\frac{d\mathbf{v}}{dt} - \frac{3R}{2}\frac{d\mathbf{u}}{dt} = -A(\mathbf{v} - \mathbf{u}) + \varepsilon\xi(t), \tag{33}$$

where $\xi(t) = [\xi_x(t),\xi_y(t)]^T$. Inertial particles are *aerosols* if $0 < R < 2/3$ and they are *bubbles* if $2/3 < R < 2$. The limit $A \to \infty$ corresponds to the situation of ideal particles (passive advection).

A convenient numerical framework to study the advective dynamics of inertial particles [11] is the open flow model of the von Kármán vortex street in the wake of a cylinder of radius r, located at $(x,y) = (0,0)$. A time-periodic stream function

$\Psi(x,y,t)$ (period $T_f = 1$ in a standard dimensionless form) governing the motions of vortices in the background flow of velocity u_0 can be constructed explicitly from the solutions of the two-dimensional viscous Navier-Stokes equations for the geometry of a circle of radius r in the middle of an infinite channel of width $w = 4r$ [25]. The Reynolds number is $R_e \approx 250$. The flow velocity $\mathbf{u}(x,y,t)$ can be obtained from $\Psi(x,y,t)$, allowing the particle motions to be computed.

In Ref. [11], it was shown that attractors can be formed in the bubble regime. It is thus convenient to focus on this regime, e.g., by fixing $R = 1.47$ and $A = 30$. There are three attractors [11]: two chaotic and one at $x = \infty$. The chaotic attractors are located near the cylinder (but not stuck on it): one in $y > 0$ and another in $y < 0$. To gain insight into what might happen to the attractors under noise, the basins of attraction of these attractors can be examined [10]. Figure 10 shows the basins of attraction of the two chaotic attractors (light blue and yellow, respectively), where the blank region denotes the basin of the attractor at infinity. Note that the phase space is five-dimensional, so what is shown in Fig. 10 is in fact a two-dimensional slice of the basin structure in the full phase space, which corresponds to the physical space. Near the cylinder, the basin boundaries among the three attractors are apparently fractal.[1] Because of the explicit time dependence in the stream function and therefore in the flow velocities, the attractors and their basins move oscillatorily around the cylinder. The remarkable feature is that in the physical space, there are time intervals during which the attractors come close to the basin boundaries. Thus, under noise, we expect permanently trapped motion on any one of the two chaotic attractors to become impossible. In particular, particles can be trapped near

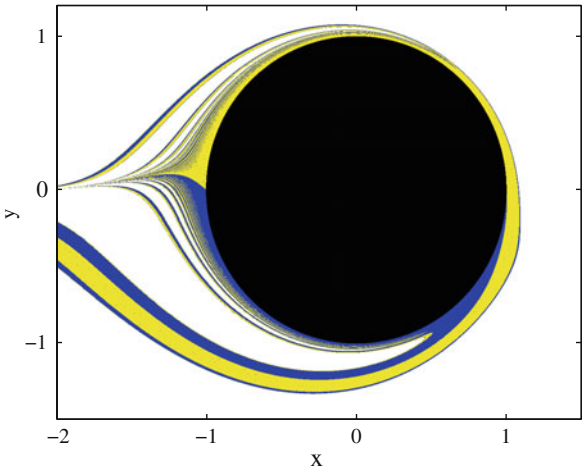


Fig. 10 Basins of attraction of two chaotic attractors (*light blue* and *yellow*, respectively) in the absence of noise. The *blank* region denotes the basin of the attractor at $x = \infty$. From [10] with permission

[1] Note that Fig. 10 represents a plot of initial conditions. For both Hamiltonian and inertial particles, those with long lifetime are close to the stable foliations of the nonattracting chaotic set. It is known that, for a general Hamiltonian system, under weak dissipation, the stable foliations are converted into the basin boundaries between the coexisting attractors [26], which are fractals.

the cylinder, switching intermittently between the two originally chaotic attractors, but this can last only for a finite amount of time: eventually all trajectories on these attractors escape and approach the $x = \infty$ attractor. That is, chaos becomes transient under noise.

To understand the nature of the noise-induced transient chaos, one can distribute a large number of particles in the original basins of the chaotic attractors and examine the channel(s) through which they escape to the $x = \infty$ attractor under noise [10]. Figure 11a, b, c show, for three instants of time (t, $t+T_f/4$, and $t+T_f/2$, respectively), locations of an ensemble of particles in the physical space. Due to the symmetry of the flow [25], the particle trajectories at t and $t+T_f/2$ are symmetric to each other with respect to the x-axis, as can be seen from Fig. 11a, c. While there are particles still trapped in the original attractors, many others are already away from the cylinder. Since this is a two-dimensional projection of a five-dimensional dynamics, some fractal-like features overlap. The channels through which they escape are a set of thin openings surrounding the cylinder and extending to one of the vortices in the

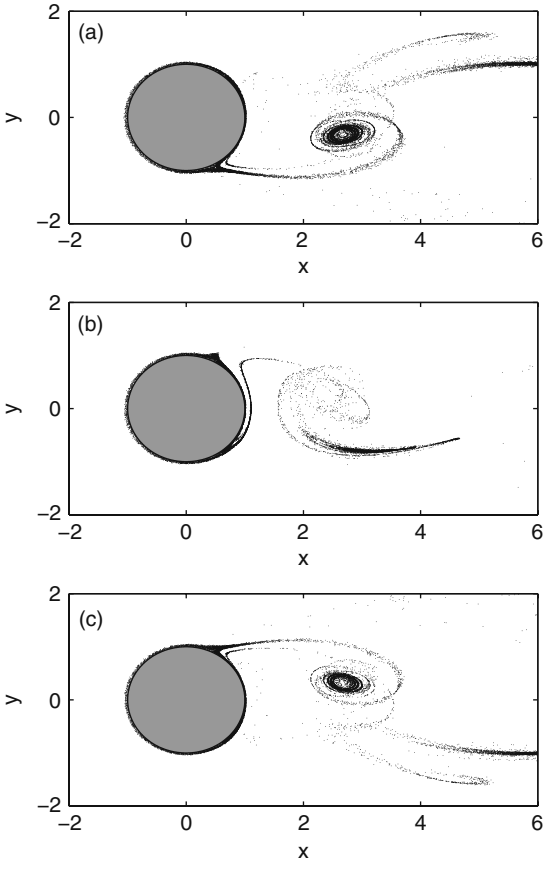


Fig. 11 (**a–c**) At three different instants of time, $T_f/4$ apart, locations of the temporally trapped and escaping particles in the physical space. From [10] with permission

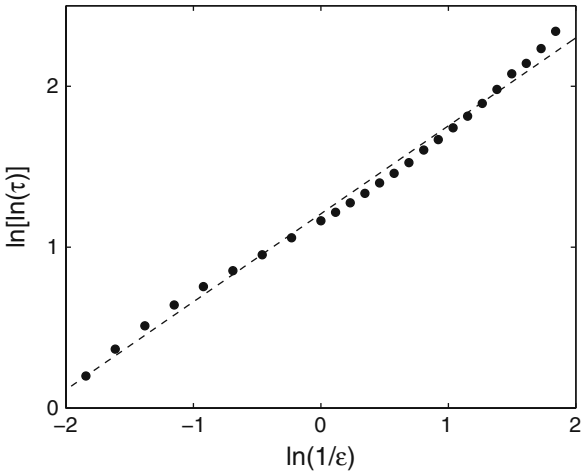


Fig. 12 Scaling of the average lifetime of the trapped chaotic particles versus the noise amplitude. From [10] with permission

flow. After wandering near the vortex, particles go to the $x = \infty$ attractor. Because of the time-dependent nature of the flow, in the physical space the locations of these openings vary in time, but the feature that they are narrow is common. For a fixed noise amplitude, numerically it was found that the lifetimes of the particles near the cylinder obey an extremely slow, exponentially decaying distribution, from which the average lifetime τ is obtained. Figure 12 shows τ versus the noise amplitude ε on a proper scale. A least-squares fit gives $\tau \approx \exp[3.3\varepsilon^{-0.55}]$. Note that for $\varepsilon = 0$, there is an attracting motion so that τ diverges. Figure 12 suggests, however, the way that τ diverges follows the superpersistent transient scaling law as ε is decreased.

Theoretically, the observed noise-induced superpersistent chaotic transient and the scaling behavior in Fig. 12 can be explained by using the approach in Sect. 5. Since the phase space in a two-dimensional fluid problem is exactly the configuration (or physical) space, the result exemplified by Fig. 12 implies that it may be possible to observe superpersistent chaotic transients in physical space. It was suggested [10] that the flow system used for experimental study of chaotic scattering [27] could be used for this purpose.

7 Conclusions

In conclusion, an unstable – unstable pair bifurcation can generate a narrow channel through which trajectories originally on a chaotic attractor can escape, converting attracting motion into a transient. The average transient lifetime depends exponentially on the time required for a trajectory to pass the channel, which in turn depends on quantities such as the parameter difference, symmetry-breaking parameter, and noise amplitude etc., typically algebraically. As a result, a superpersistent chaotic transient arises. The transients can accompany phenomenon such as the onset of riddled basins and the stability of attractors formed by inertial particles advected

in hydrodynamical fluid flows. Such transients are also expected to be common in spatially extended dynamical systems.[2]

Acknowledgement This work was supported by AFOSR under Grant No. F9550-09-1-0260.

References

1. C. Grebogi, E. Ott, and J. A. Yorke. Chaotic attractors in crisis. *Phys. Rev. Lett*., **48**, 1507–1510 (1982).
2. C. Grebogi, E. Ott, and J. A. Yorke. Crises, sudden changes in chaotic attractors and chaotic transients. *Phys. D*, **7**, 181–200 (1983).
3. C. Grebogi, E. Ott, and J. A. Yorke. Fractal basin boundaries, long-lived chaotic transients, and unstable-unstable pair bifurcation. *Phys. Rev. Lett*., **50**, 935–938 (1983).
4. C. Grebogi, E. Ott, and J. A. Yorke. Super persistent chaotic transients. *Ergod. Th. Dym. Syst*., **5**, 341–372 (1985).
5. Y.-C. Lai, C. Grebogi, J. A. Yorke, and S. Venkataramani. Riddling bifurcation in chaotic dynamical systems. *Phys. Rev. Lett*., **77**, 55–58 (1996).
6. J. C. Alexander, J. A. Yorke, Z. You, and I. Kan. Riddled basins. *Int. J. Bifur. Chaos Appl. Sci. Eng*., **2**, 795–813 (1992).
7. J. R. Crutchfield and K. Kaneko. Are attractors relevant to turbulence? *Phys. Rev. Lett*., **60**, 2715–2718 (1988).
8. V. Andrade, R. Davidchack, and Y.-C. Lai. Noise scaling of phase synchronization of chaos. *Phys. Rev. E*, **61**, 3230–3233 (2000).
9. M. G. Rosenblum, A. S. Pikovsky, and J. Kurths. Phase synchronization of chaotic oscillators. *Phys. Rev. Lett*., **76**, 1804–1807 (1996).
10. Y. Do and Y.-C. Lai. Superpersistent chaotic transients in physical space: advective dynamics of inertial particles in open chaotic flows under noise. *Phys. Rev. Lett*., **91**, 224101 (2003).
11. I. J. Benczik, Z. Toroczkai, and T. Tél. Selective sensitivity of open chaotic flows on inertial tracer advection: catching particles with a stick. *Phys. Rev. Lett*., **89**, 164501 (2002).
12. B. R. Hunt and E. Ott. Optimal periodic orbits of chaotic systems occur at low period. *Phys. Rev. E*, **54**, 328–337 (1996).
13. M. V. Jacobson. Absolutely continuous measures for one-parameter families of one-dimensional maps. *Commun. Math. Phys*., **81**, 39 (1981).
14. U. Frisch. *Turbulence*, 1st ed. Cambridge University Press, Cambridge, UK (1995).
15. I. Tsuda. Self-similarity in the belousov-zhabotinsky reaction. *Phys. Lett. A*, **85**, 4–8 (1981).
16. Y. Do and Y.-C. Lai. Extraordinarily superpersistent chaotic transients. *Europhys. Lett*., **67**, 914–920 (2004).
17. Y. Do and Y.-C. Lai. Scaling laws for noise-induced superpersistent chaotic transients. *Phys. Rev. E*, **71**, 046208 (2005).
18. C. W. Gardiner. *Handbook of Stochastic Methods*, 1st ed. Springer-Verlag, New York (1997).
19. H. Risken. *The Fokker-Planck Equation*, 1st ed. Springer-Verlag, Berlin (1989).
20. H. Aref. Stirring by chaotic advection. *J. Fluid Mech*., **143**, 1–21 (1984).
21. J. M. Ottino. *The Kinematics of Mixing: Stretching, Chaos and Transport*, 1st ed. Cambridge University Press, Cambridge, UK (1989).
22. M. R. Maxey and J. J. Riley. Equation of motion for a small rigid sphere in a nonuniform flow. *Phys. Fluids*, **26**, 883–889 (1983).
23. J. Rubin, C. K. R. T. Jones, and M. Maxey. Settling and asymptotic motion of aerosol-particles in a cellular-flow field. *J. Nonlinear Sci*., **5**, 337–358 (1995).

[2] For a recent review on this topic, see [28].

24. T. J. Burns, R. W. Davis, and E. F. Moore. A perturbation study of particle dynamics in a plane wake flow. *J. Fluid Mech*., **384**, 1–26 (1999).
25. C. Jung, T. Tél, and E. Ziemniak. Application of scattering chaos to particle transport in a hydrodynamical flow. *Chaos*, **3**, 555–568 (1993).
26. A. E. Motter and Y.-C. Lai. Dissipative chaotic scattering. *Phys. Rev. E*, **65**, 015205 (1–4) (2002).
27. J. C. Sommerer, H.-C. Ku, and H. E. Gilreath. Experimental evidence for chaotic scattering in a fluid wake. *Phys. Rev. Lett*., **77**, 5055–5058 (1996).
28. T. Tél and Y.-C. Lai. Chaotic transients in spatially extended systems. *Phys. Rep*. **460**, 245–275 (2008).

Synchronization in Climate Dynamics and Other Extended Systems

Peter L. Read and Alfonso A. Castrejón-Pita

Abstract Synchronization is now well established as representing coherent behaviour between two or more otherwise autonomous nonlinear systems subject to some degree of coupling. Such behaviour has mainly been studied to date, however, in relatively low-dimensional discrete systems or networks. But the possibility of similar kinds of behaviour in continuous or extended spatiotemporal systems has many potential practical implications, especially in various areas of geophysics. We review here a range of cyclically varying phenomena within the Earth's climate system for which there may be some evidence or indication of the possibility of synchronized behaviour, albeit perhaps imperfect or highly intermittent. The exploitation of this approach is still at a relatively early stage within climate science and dynamics, in which the climate system is regarded as a hierarchy of many coupled sub-systems with complex nonlinear feedbacks and forcings. The possibility of synchronization between climate oscillations (global or local) and a predictable external forcing raises important questions of how models of such phenomena can be validated and verified, since the resulting response may be relatively insensitive to the details of the model being synchronized. The use of laboratory analogues may therefore have an important role to play in the study of natural systems that can only be observed and for which controlled experiments are impossible. We go on to demonstrate that synchronization can be observed in the laboratory, even in weakly coupled fluid dynamical systems that may serve as direct analogues of the behaviour of major components of the Earth's climate system. The potential implications and observability of these effects in the long-term climate variability of the Earth is further discussed.

P.L. Read (✉)
Atmospheric, Oceanic & Planetary Physics, Clarendon Laboratory, Department of Physics, University of Oxford, Oxford, OX1 3PU, UK
e-mail: p.read1@physics.ox.ac.uk

M. Thiel et al. (eds.), *Nonlinear Dynamics and Chaos: Advances and Perspectives,* Understanding Complex Systems, DOI 10.1007/978-3-642-04629-2_8,

1 Introduction

In recent years, the study of synchronization phenomena in nonlinear systems has made a number of significant advances in various areas of physics, engineering and the life sciences [1]. The first documented example of synchronization was reported as long ago as 1665 by Christiaan Huygens, who noted the tendency of a pair of pendulum clocks, mounted on a common support, eventually to synchronize the phase of their oscillations, even if they would adopt slightly different swing periods if isolated from each other. The two pendula were "coupled" by the common support which was moving in a way which was hardly perceptible to Huygen's eyes but strong enough for the pendula to "feel" each other, i.e. the two pendula were effectively interacting through small vibrations in the support. Later on, it was concluded that these small vibrations produced the synchronization of the two pendula.

Synchronization is a phenomenon that nowadays is usually treated as a regime in which two, or more, coupled periodic or even chaotic systems exhibit correlated, and sometimes even identical, oscillations. A key feature is that at least one of the systems is sufficiently nonlinear that external perturbations can affect its natural oscillation frequency and hence its phase. The applied coupling is therefore able to systematically modify the behaviour of the perturbed system to result in a regime in which its phase of oscillation is locked into a coherent relationship with that of the perturbing system (and vice versa in the case of two-way coupling).

Since the first investigations of synchronization phenomena, several different types of synchronization have been identified and studied. These include situations where the locking of both the phase *and amplitude* of the systems may be less than perfect. Such partially synchronized states include phase synchronization, lag synchronization, complete synchronization, intermittent and imperfect (phase) synchronization, and so-called generalized synchronization [1, 3]. It has been suggested that all these different synchronization states are in fact related to each other and are different manifestations of the same universal phenomenon. However, the relationship between them is still not well understood or clarified [4].

In this paper we examine aspects of the climate of the Earth to which ideas related to synchronization might apply. Much effort is put into trying to understand the complexities of the climate system and its dynamical evolution by the scientific community, and to develop models for predicting its behaviour and response to various perturbations, both natural and anthropogenic. The possible tendency for certain sub-systems within such a complex dynamical network to develop even partially synchronized behaviour could have major implications for its future evolution. So it is of increasing importance for nonlinear dynamical theory to play a full role in advancing our understanding of the impact of both transient and continuous perturbations on systems as complex as the climate. Section 2 therefore reviews some aspects of climate variability that can be viewed as representing quasi-periodic or chaotic oscillations, potentially subject to synchronization with external or internal perturbations. In Sect. 3 we further develop these ideas in the context of models of so-called teleconnections, representing various forms of correlated behaviour observed in the climate between regions that are geographically distant. Section 4

goes on to consider some simple model systems that may provide some useful insights into a new class of possible synchronized states in extended systems analogous to the climate, coupled via the large-scale axisymmetric components of their circulation. Finally, we offer some concluding remarks and an outlook for further research in Sect. 5.

2 Climate Cycles and Teleconnections

The origin of oscillations and cyclic behaviour in the weather and circulation of the atmosphere and oceans has long been of interest to scientists, as well as to those involved in weather-sensitive industries such as agriculture and fisheries. These roughly cyclic oscillations typically vary on time scales from a few weeks to decades and even millennia and longer intervals. They may also either occur across the globe, or focussed in particular geographical areas. In the latter case, however, the influence of a localised disturbance may be still detectable in remote locations. This is known in the climatological literature as a "teleconnection". We examine examples of these types of phenomenon in this section, with particular reference to situations where nonlinear synchronization may play a role.

2.1 Cyclic Variations in Climate Variables

On some of the longest timescales amenable to quantitative study, the onset and retreat of major glaciations is perhaps the best known example of a cyclic climate phenomenon. Information gleaned from deep ice cores at both polar ice caps show that the volume of the Earth's polar ice sheets has varied considerably during the past million years or so from its present configuration, with advancing and retreating glaciation of the mid-latitude regions of both hemispheres taking place with a repeat period of around 100 kyrs. These cyclic glaciations appear to be a comparatively recent phenomenon in geological terms, associated with a gradual cooling of the entire planet since the relatively warm Eocene period, some 55 Myrs ago [5]. A particularly notable feature of these recent glaciation cycles, however, is their apparent relationship to cyclic variations in the Earth's orbit and rotation, known as the *Milankovitch cycles* after their Serbian discoverer in the early twentieth century. Detailed analyses have indicated a degree of apparent synchronization in the phase of the advance and retreat of glaciations with certain components of the Milankovitch cycles, which mostly include periods of between 18 and 41 kyrs. However, the degree of synchronization continues to be controversial [6], not least because the glaciations are dominated by a roughly 100 kyr period which lies outside the main band of orbital and rotational periodicities. It remains unclear, therefore, whether the Milankovitch orbital/rotation cycles *force* the observed glaciations or merely act as the "pacemaker" [7] of oscillations that would take place with or without such external forcing.

On much shorter timescales, commonly cited examples of cyclic behaviour in the climate system include the El Niño/Southern Oscillation (ENSO), the Quasi-biennial Oscillation (QBO) in the Earth's stratosphere, the North Atlantic Oscillation/Arctic Oscillation (NAO/AO) and the so-called "zonal index cycle". ENSO is perhaps the best known of the interannual climate cycles in the troposphere. It is related to the extent of an accumulation of relatively cool water close to the ocean surface off the coast of equatorial South America, the presence of which is maintained by coupled circulation patterns in both the Pacific Ocean and the overlying atmosphere. Fluctuations in this coupled circulation leads to occasional weakening, which allows the cold upwelling in the oceans to spread westwards and cause the tropical ocean temperature across the equatorial Pacific to cool significantly. Such episodes (known as El Niño events) recur on timescales of 3–6 years, which appear to be mainly determined by the time it takes equatorially trapped Kelvin and Rossby waves in the ocean to cross the Pacific. The shift in ocean surface temperatures then lead to significant changes in seasonal climate at many locations across the globe. A particular feature of El Niño events is their tendency to occur at particular times of year (particularly around December – January, hence the name "El Niño" meaning the Christ child), suggesting at least a partial synchronization with the seasonal cycle. The basic mechanism that sets the timescale of the Southern Oscillation, however, is not directly affected by the seasonal cycle. But this appears to be another case where an astronomical cycle acts as a partial "pacemaker", modifying the phase of particular features within an otherwise free (though almost certainly chaotic) oscillation.

The stratospheric QBO is another example, in which the annual cycle plays the role of a partial pacemaker. The oscillation manifests itself as a cyclic reversal of the prevailing east – west wind in the tropics at altitudes above 10–15 km above the surface, with a recurrence period of around 24–29 months. It is now widely understood to be a form of relaxation oscillation, driven by the rectification of upward-propagating waves originating in the convectively active troposphere [8], with the timescale set by the mean intensity of the convective forcing. This appears to be a ubiquitous form of oscillation in planetary atmospheres, since similar types of behaviour have now been identified in the stratospheres of Jupiter [9], Saturn [10] and even Venus [11], though on somewhat different timescales. But for the Earth the seasonal cycle itself modulates this convective forcing, allowing the possibility for nonlinear phase-locking effects to modify the onset of wind reversals, which tend to take place at preferred times of the year [12].

The so-called "index cycle" is among the shortest timescale cyclic phenomena in the climate system, being representative of a class of phenomena known as "intraseasonal oscillations". This was first noted in the northern mid-latitude circulation over the Atlantic Ocean [13], and appears as a tendency for the main east – west jet stream in the upper troposphere (at altitudes of around 10 km) to alternate between a fast, relatively undisturbed state and a more undulating, "blocked" configuration, in which travelling waves are of much larger amplitude but relatively slowly moving. This alternation occurs roughly cyclically at certain times of year on timescales of around 30–50 days, though because of the disturbances to the circulation associated with the contrast between continents and oceans it is often

difficult to discern. The term "index cycle" refers to a statistical index – the "zonal index", defined as a measure of the average east-west (zonal) wind between 35° and 55°N – from which timeseries could be constructed that showed evidence for cyclic behaviour, though this often appeared to be highly chaotic. There is a similar phenomenon discernible also in the southern hemisphere, where continentality has a much weaker influence on the circulation than in the north. Although the latter has been rather less intensively studied, it does reveal this type of cyclic behaviour more clearly (see Fig. 1) than in the northern hemisphere [14]. This is of particular interest, because it appears to represent a nonlinear oscillation associated with cyclic exchanges of energy (potential and kinetic) between travelling baroclinic cyclone waves and the zonally-symmetric component of the circulation. Such wave-zonal flow interactions are the dominant form of nonlinear interactions involving large-scale atmospheric waves, a point we return to in Sect. 3.2.

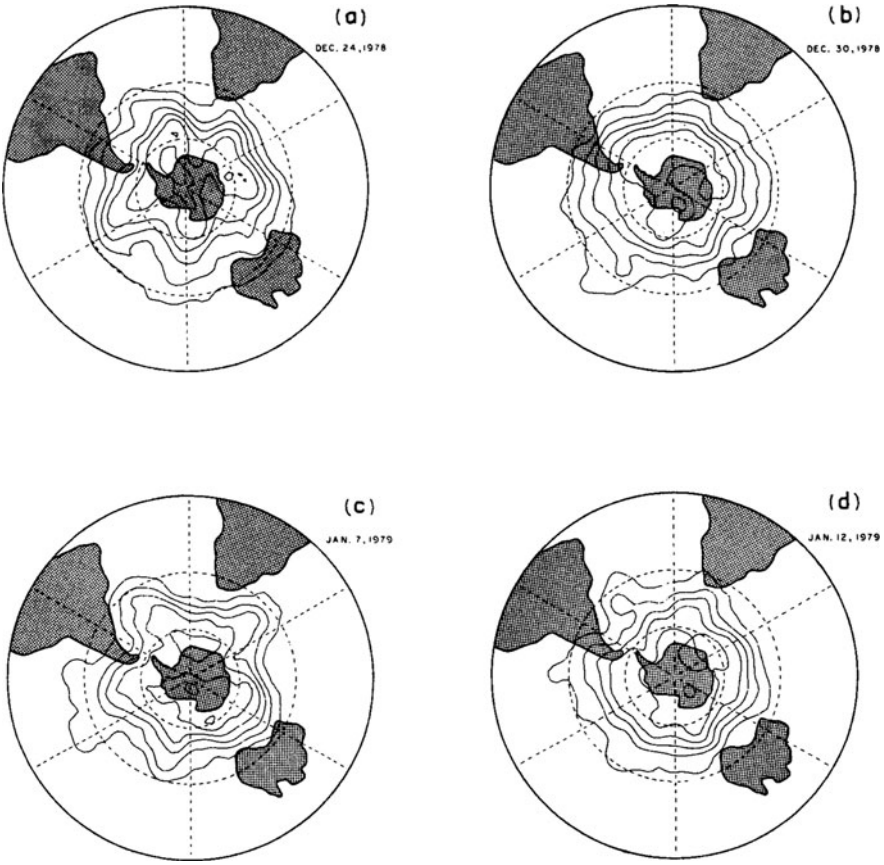


Fig. 1 Sequence of 200 mb geopotential height contours over the southern hemisphere during the summer of 1979, illustrating the alternating of the circulation between "westerly" and "blocked" states during the southern "index cycle". Figure adapted from [14]

2.2 Teleconnections

Cyclic phenomena such as the glaciation cycles appear to manifest themselves on a global scale, so can be regarded as oscillations of the entire climate system virtually worldwide. Other types of climate cycles and variations, however, appear to be more confined to particular geographical regions, although their influence may be felt further afield in other parts of the world.

The North Atlantic Oscillation is one example of such a regional phenomenon that has been studied for almost 200 years [15]. It appears in the form of statistically correlated fluctuations in time between widely separated points spanning the North Atlantic basin, commonly applied to variability on monthly and longer timescales. Such long term, correlated fluctuations are known as "teleconnections", though the origin and mechanisms for these phenomena are often difficult to determine. But with the development of the large-scale global observing network for weather prediction and climate studies, the study of teleconnection phenomena has revealed a number of recurrent and persistent patterns that vary on timescales ranging from the intraseasonal (30–80 days) to the interdecadal. In the case of the NAO, this appears to comprise a combination of fluctuations that include the zonal index cycle mentioned above, on intraseasonal timescales and with a spatial distribution that is largely symmetric about the pole (a so-called "annular mode") and a more complex pattern of correlation that recurs on interannual timescales [16].

Such complex spatial patterns of correlation often appear to take on a character akin to a standing wave pattern. This has led to a widely accepted paradigm within climate science of a quasi-linear mechanism for teleconnections that involves the excitation and propagation of large-scale Rossby waves from a specific source region. The source might consist, for example, of an unusually warm patch of ocean surface, brought about by anomalies in the ocean circulation which, if persistent, may lead to a standing wave pattern aligned along great-circle paths across the globe [17]. The simplest case under this scenario is that the temporal correlation of atmospheric variables between two points in space is due to the fact that these two locations are along the "track" of a major atmospheric anomaly.

Another widely discussed interpretation within meteorology and climate for exciting these kinds of correlated fluctuations is to regard them as an aggregation of quasi-stochastic weather events through either a physical or statistical low-pass filter [18]. Such an approach largely involves the application of linear dynamics, much like the effect of noise on a resonant oscillatory system.

There is, however, one further type of teleconnection that has been identified recently and that is not easily explained in terms of linear dynamics. This involves the identification of phase-synchronized behaviour of regional weather events, that would seem to imply a nonlinear phase-locking mechanism between geographically separated meteorological phenomena without recourse to any external forcing. Teleconnections between middle latitude blocking of the Northern and Southern hemispheres, for example, have been diagnosed in climate data [19], manifested as a small but significant tendency for such blocking to occur simultaneously in both the northern and southern hemispheres. This was found to be the case, even though

the events considered singly in each hemisphere appear to be chaotic in time. The blocked states essentially represent a particular phase in the respective zonal index cycles of the two hemispheres, as introduced in the previous subsection. However, the mechanism for producing such correlated behaviour between two hemispheres, often at quite different points in their respective seasonal cycles, is far from clear.

3 Models and Mechanisms for Teleconnection and Synchronization

In the previous section we have introduced a number of phenomena within the Earth's climate system in which cyclic, though usually chaotic, oscillatory behaviour is observed. The origin of such cyclic behaviour is often difficult to determine, however, and so it is helpful to appeal to simplified models in order to evaluate possible mechanisms. Where synchronized behaviour between two or more oscillatory phenomena is observed, then two distinct questions arise: (a) does the oscillation in one system owe its sole origin to that of the other? or (b) are the two cyclic oscillations sustainable independently but simply brought into a synchronized state through some form of weak coupling? If the latter, how weak and/or intermittent can this coupling be in order to exhibit observable synchronization?

3.1 Distinguishing Synchronized Models from Observations?

Where the perturbation is a highly predictable external process, such as the astronomical orbit/rotation cycles invoked for the Milankovitch cycles of glaciation, the forcing or synchronization is clearly a unidirectional problem, in which the external forcing influences the climate system and not vice versa. In this case, it remains controversial as to whether the glaciation cycles would continue in the absence of cyclic variations in orbit and rotational parameters. Because of the very large timescales involved, it is not feasible to make use of comprehensive physical climate models to investigate this problem. Instead, it is necessary to appeal to much simpler models, based typically on energy balance considerations for the land, oceans and cryosphere to represent the integrated response of the Earth on such long timescales. Such low-order models may be very crude in comparison to the real Earth system. Even so, various authors have sought to draw conclusions from studies using low-order energy-balance models in which the actual variation in solar heating at high latitudes due to orbital forcing is applied. An example of such a study is illustrated in Fig. 2, which shows the result of applying various forms of weak orbital forcing to a very simple nonlinear energy balance climate model [20].

A major difficulty has emerged, however, in that a wide variety of models appear able to come into synchronization with the applied orbital forcing [20] (see Fig. 2) with little apparent sensitivity to the details of the model itself. The resulting "prediction" for ice volume variations then seems able to reproduce the observed variations, provided only that the representation of glaciation in the model is nonlinear.

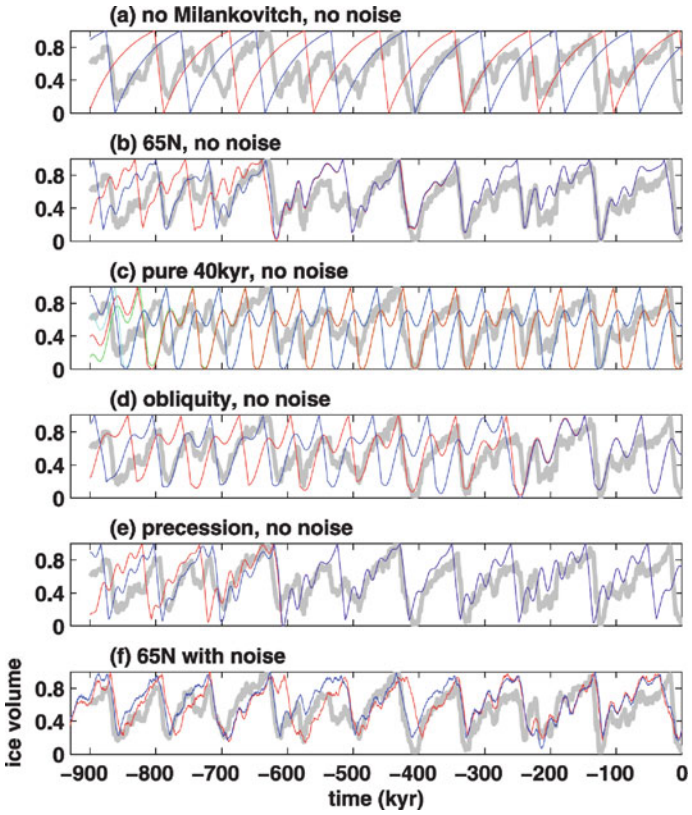


Fig. 2 An illustration of phase locking to Milankovitch forcing. The *gray curve* is a timeseries of the proxy δ^{18}O record representing the observed variations of global temperature; the *thin color curves* are ice volume time series from different model runs by Tziperman et al. [20] using different initial conditions. (**a**) A model run with no Milankovitch forcing. (**b**) Model forced by and phase locked to Milankovitch forcing (65°N summer insolation). (**c**) Model run forced by and phase locked to an artificial, periodic 40 kyr forcing. (**d**) Model forced by obliquity variations only. (**e**) Model forced by precession only. (**f**) Same as (**b**) but in the presence of noise. Figure from [20] with permission

Though clearly encouraging as an indication that even weak Milankovitch perturbations in insolation can lead to synchronized glaciation cycles, this result also demonstrates that simple comparisons between forcing input and model predictions are not sufficient to distinguish between even major classes of models very effectively—a clear consequence of synchronization within a dissipative nonlinear system.

The situation where synchronization occurs between two autonomous, chaotic components of the climate system may therefore be even more difficult and subtle to elucidate. In the synchronized blocking case introduced above, and also in a related study [21] that considers coupling between nearby longitudinal sectors in the same hemisphere, attempts to model this phenomenon have made use of simplified numerical models of flow in a hemispheric channel. Each hemisphere was

represented as a separate model, and coupling between the hemispheres or sectors was then introduced by adding a heuristic diffusive coupling between pairs of spatial harmonics. For their study, however, Duane & Tribbia [21, 22] chose to couple only selected pairs of wavenumber modes that explicitly excluded $|K| < k_0$, where K is the total wavenumber $K^2 = k^2 + l^2$, k_0 is a constant and k and l are the respective wavenumbers in the zonal and meridional directions. Such a restricted coupling reflects the common teleconnection paradigm discussed above, which postulates trains of Rossby-Haurwitz waves as the conduit for dynamical information across the globe, although there is as yet no direct evidence for this mechanism in observations for this mode of coupling in coupled blocking events. Nevertheless, the models thus formulated were able to show evidence for correlated (or, in the case of two sectors, anti-correlated) blocking events between hemispheres or sectors in a manner that was at least qualitatively comparable to the observations. However, in light of the remarks given above concerning the Milankovitch forcing, an ability of a given model to reproduce synchronized behaviour does not guarantee that it represents the correct physics of the basic oscillation or coupling without further evidence.

3.2 Zonally Symmetric Coupling

Motivated in part by studies over many years of quasi-periodic and chaotic "index cycles" in laboratory experiments [23, 24], we have investigated synchronization effects in a baroclinically unstable channel in which we make the opposite assumption to [22] and [21], coupling only via the lowest wavenumber zonally-symmetric components of the flow [25]. Like the study of [21], this work relies on the use of a quasigeostrophic (QG) two-layer formulation of baroclinic and/or barotropic instability. Since the first studies of baroclinic waves in the atmosphere, the two-layer model has played an important role. It was proposed more than 50 years ago to encapsulate in the simplest way some of the principal features of the middle latitude atmospheric circulation. Since then, it has been widely used in studies of baroclinic instability [26–29]. The basic model is a two-layer quasi-geostrophic system, formulated in Cartesian geometry, and permits just one zonally propagating wave with barotropic and baroclinic components. Hence it is arguably the simplest possible representation of such a system.

A schematic diagram of the two-layer system is shown in Fig. 3, commonly known as the "Phillips-Pedlosky" two-layer model [26–29]. It consists of two superposed fluids confined to a rotating, rectangular, and zonally periodic channel. By convention, layer 1 is at the top and layer 2 is at the bottom. The height of the channel is D, with width L. It was developed using Cartesian co-ordinates, where x, y and z are used respectively to denote displacements along the channel, across the channel, and vertically. Periodic boundary conditions are applied in x. The corresponding velocities are u_n,v_n and w_n, where $n = 1,2$ refers to the upper and lower layer respectively. The densities of the upper and lower layer are ρ_1 and ρ_2, with $\rho_2 > \rho_1$. The analysis carried out by [30, 31] was conducted in a reference frame that moved with the zonal velocity, $(u_1 + u_2)/2$, of the mean flow. Therefore, both

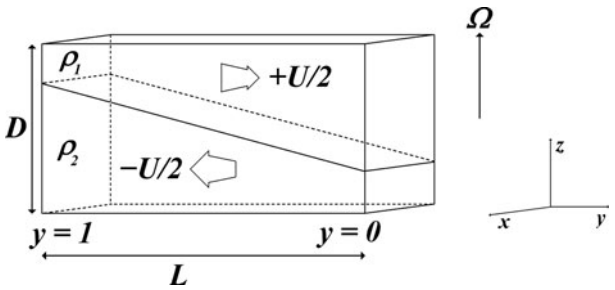


Fig. 3 The two-layer model considered in this project: two immiscible fluids of densities ρ_1 and ρ_2 with $\rho_1 < \rho_2$. The fluids are in relative motion, with velocities of $+U/2$ and $-U/2$ respectively. The channel has height D, width L and is rotating at a constant angular velocity Ω

the upper and lower layers appear to be in uniform motion along the channel in this frame of reference, with respective velocities of $+U/2$ and $-U/2$, becoming $+1/2$ and $-1/2$ in non-dimensional terms. In the absence of motion, the fluid layers are assumed here to have equal depths of $D/2$. To simulate the experimental conditions, the fluid is confined vertically between two rigid planes representing a base and a lid. The kinematic viscosity within both layers is assumed to be equal.

For maximum simplicity, a 5-dimensional version of the model was considered [30, 31], representing a spectrally highly truncated formulation that describes the interaction of a single baroclinic wave and a zonal flow. The basic 5-dimensional model equations can also be represented in complex form as a 3-equation formulation (1), in which the first equation in the complex variable X describes a barotropic travelling wave component, the next equation in Y (also complex) represents the baroclinic travelling wave (with the same zonal wavenumber) and the third component Z represents the (real) amplitude of the baroclinic zonal flow component. The resulting system of equations is almost identical to the well known 3-component complex Lorenz equations [31].

$$\begin{aligned} \dot{X} &= -\sigma X + \sigma Y, \\ \dot{Y} &= RX - aY - XZ, \\ \dot{Z} &= -bZ + \frac{1}{2}\left(X^* Y + X Y^*\right), \end{aligned} \tag{1}$$

where σ and b are real numbers signifying the Prandtl number of the fluid, and the aspect ratio of the system respectively, and R (the "Rayleigh number") and a are complex parameters (in contrast to the original complex Lorenz equations).

In the study by [25], an external periodic forcing was added (via the baroclinic zonal velocity component) to this model in order to allow the generation of various degrees of synchronized states, depending on the frequency detuning between the natural amplitude modulation oscillation of the system and the forcing.

Investigations were carried out in the three principal flow regimes that the model exhibited, a steady amplitude wave regime, an amplitude-modulated (quasi-periodic amplitude vacillation or AV) regime and a chaotic amplitude-modulated regime.

In the steady regime, amplitude resonance at certain forcing frequencies was found, in which the initial unperturbed travelling wave solution grew in amplitude as certain frequencies in the perturbation were approached. In the AV regime with relatively small amplitudes of forcing, classical "Arnol'd tongues" and synchronization were found, in which the natural frequency of amplitude modulation within the model became phase-locked to that of the external perturbation. This implies that the amplitude variations in the travelling wave became entrained with the zonally symmetric forcing. Such entrainment was only possible because of the strongly nonlinear wave-zonal flow interaction that underlies the basic instability that sustains the growth of the travelling wave. The translation speed and frequency of the travelling wave were not directly modulated by the external perturbation, but were affected indirectly by the perturbation though not explicitly phase-locked. With moderate forcing amplitudes in this regime, period doubling cascades were found, leading ultimately to an induced or enhanced chaotic state.

An interesting aspect of these results was the clear ability of the system to phase lock to harmonics of the applied perturbation frequency. Thus, in Figs. 4 and 5 we see evidence of locked states in frequency ratios corresponding to period 2, 3 and so on up to period 8, although the frequency width of such high order synchronized states is then very narrow.

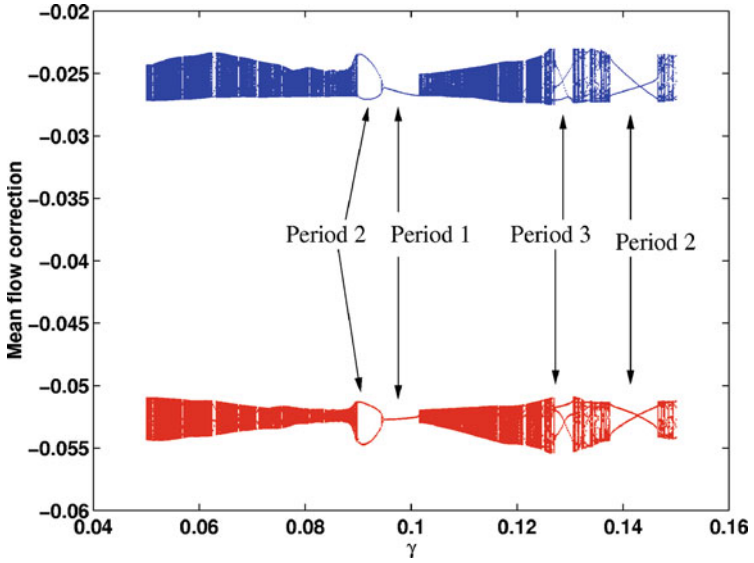


Fig. 4 Maximum and minimum zonal flow amplitudes, $X_{\min}$ and $X_{\max}$, as a function of perturbation frequency γ for forcing amplitude $\epsilon = 0.005$ for the amplitude vacillation regime in the periodically perturbed model of [25]

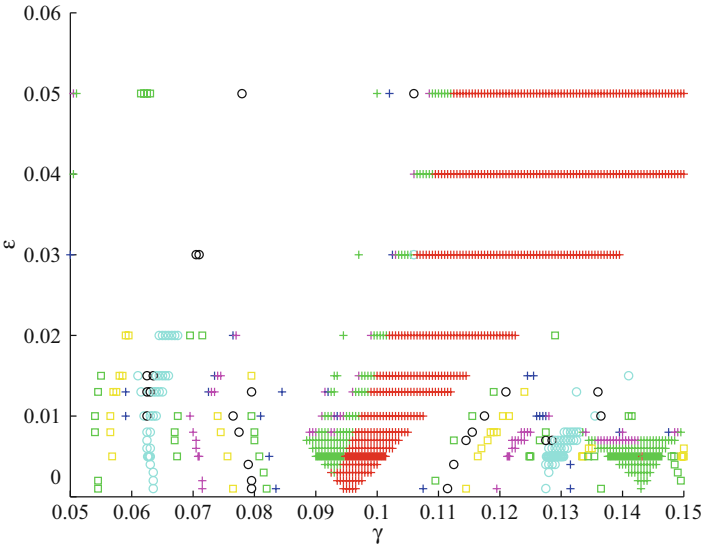


Fig. 5 The amplitude vacillation regime: plot of frequency-locked regions in the periodically perturbed model of Eccles et al. [25]. Period 1 +, period 2 +, period 3 o, period 4 +, period 5 o, period 6 o, period 7 □, period 8 +. Note the fields have not been contoured as they are highly complex and contouring may give a misleading impression

In contrast, when the system was run in a chaotic, amplitude-modulated regime (when unperturbed), the introduction of periodic forcing was able to suppress this chaotic behavior at particular regions in the parameter space, via a form of open-loop chaos control. Thus, despite employing a very restricted form of coupling that only perturbs the zonally symmetric component of the flow, there was clearly enough nonlinearity and dissipation within the system to exhibit a fairly full range of partial and complete synchronization phenomena.

More recently, the approach of [25] has been extended to allow for the coupling of two non-identical 5-dimensional baroclinic channel models, effectively representing the zonally-symmetric coupling of two hemispheres or zonal sectors [32].

The interaction between the two models was achieved by introducing a linear coupling term in the *response* or *slave* system, in the case of a *master-slave* configuration or in both models in the case of mutual or bidirectional configuration. This coupling term is proportional to the difference between the different variables of the system of the form:

$$\frac{d\mathbf{X_1}}{dt} = G(\mathbf{X_1}) + \eta_1 \mathbf{E}(\mathbf{X_2} - \mathbf{X_1}), \tag{2a}$$

$$\frac{d\mathbf{X_2}}{dt} = G(\mathbf{X_2}) + \eta_2 \mathbf{E}(\mathbf{X_1} - \mathbf{X_2}). \tag{2b}$$

where $\mathbf{E}$ is the matrix that determines the linear combination of the components of $\mathbf{X}_{1,2}$ that will be used to couple the systems, and η_1 and η_2 are the coupling

strengths. If $\eta_2 = 0$ (the same applies for $\eta_1 = 0$) a typical *master-slave* configuration is obtained. On the other hand, if $\eta_2 = \eta_1$ a symmetrical *mutual* coupling is achieved. So far, we have only used these two configurations. The de-tuning between the systems was produced by having different values for the rotational Froude number, $F_{1,2}$.

By choosing

$$\mathbf{E} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and taking into consideration the 3-equation formulation of our uncoupled systems (1) we introduce a zonally symmetric coupling on the Z component of the system. Although this approach is formally ad hoc, it is noteworthy that it has the same form as Newton's law of cooling $dT/dt = K(T_0 - T)$, which is appropriate for forced convective heat transfer. The latter is essentially how the thermal boundary conditions are maintained in typical baroclinic annulus experiments. This is therefore, a form of coupling that will directly link the amplitudes of the baroclinic zonally symmetric flow components of the two systems.

Investigations using the *master-slave* configuration ($\eta_1 = 0$) [32] have demonstrated the ability of this simple system to exhibit various degrees of synchronization in both periodic and chaotic states. It was also found that the frequency range over which synchronized states were found, for low and moderate coupling strengths, increased with increasing η, following an Arnol'd tongue-like behaviour. Larger coupling strengths led to period-doubling and higher order (2:1) phase synchronization.

This system has also proved fruitful for the application of the *auxiliary system approach* [33] as a means of seeking evidence for generalised synchronization, in the form of yet another degree of synchronization. It was generally found that larger values of η are needed to cross the boundary from phase synchronized states to fully synchronized states in the generalised sense. It was also found that larger coupling strengths were needed to obtain phase synchronization in chaotic regimes than in periodic ones. Some investigations were also carried out by coupling the system in mixed states, i.e. the *master* system was run in a region in parameter space where the dynamics was periodic, while the *slave* system was in a chaotic one. This can be interpreted as a direct extension of the periodically forced system mentioned above and described by Eccles et al. [25]. Likewise, phase synchronization and signs of chaos control and chaos-destroying synchronization were also found.

For the bidirectional coupling case ($\eta_1 = \eta_2$), signatures of mutual synchronization such as imperfect phase synchronization (Fig. 6) and complete phase synchronization were detected [32] in both periodic and chaotic regimes. Complex effects such as oscillation quenching were also found, which do not have an equivalent in the unidirectionally coupled systems. In this case, in contrast to the previous scenario of unidirectional coupling, both sub-systems are affected by the coupling, and the observed amplitude modulation frequencies of the two systems were brought

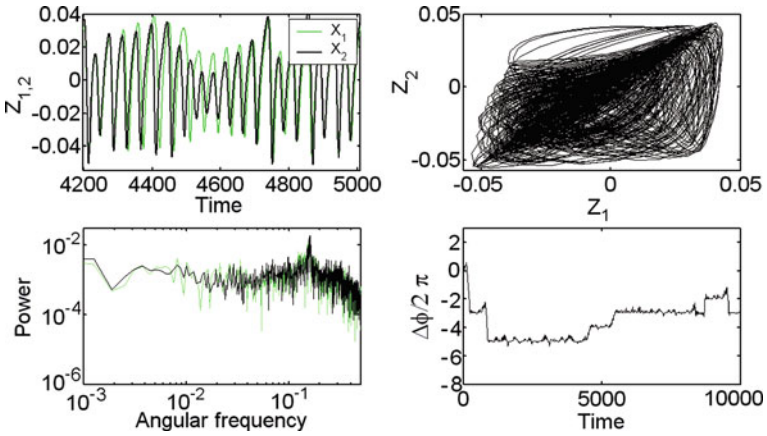


Fig. 6 Behaviour of a mutually (symmetric) coupled system in a chaotic state, for which $F_a = 13$, $F_b = 13.55$, and $\eta = 0.02$. (**a**) Time series for system 1 (*black line*) and system 2 (*green line*); (**b**) Lissajous figure from plotting $Z_a(t)$ against $Z_b(t)$; (**c**) Frequency spectra for system 1 (*black line*) and system 2 (*green line*); (**d**) Phase difference $\Delta\phi(t)$ between Z of the system 1 and system 2. Notice that system 2 loses one cycle against system 1 at around Time= 4,600 in (**a**) and its correspondent phase slip in (**d**); this is a clear signature of imperfect phase synchronization

together to oscillate at a common frequency. In this configuration, it was found that the synchronized regions in the coupling-strength/de-tuning diagram (shown in Fig. 7 for the chaotic case) are quite symmetrical. It is also observed that relatively large coupling strengths are needed to reach full phase synchronization for relatively

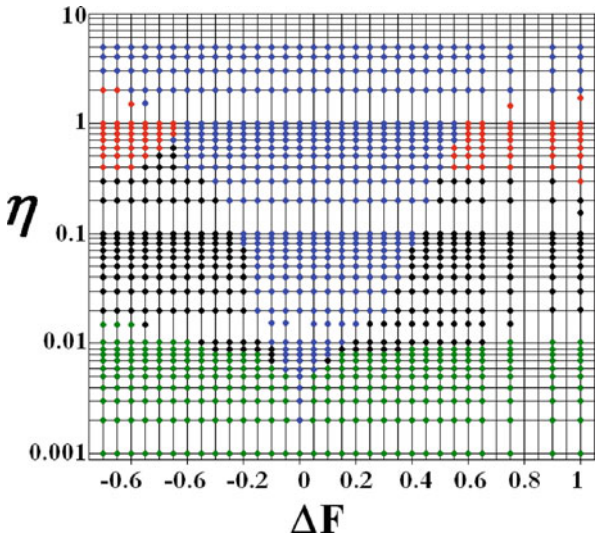


Fig. 7 The chaotic modulated amplitude vacillation regime: diagram of synchronized regions in the mutually coupled model developed in [32]. Unsynchronized ●, imperfect phase synchronized (phase slips) ●, phase synchronized ●, oscillation quenching ●

small de-tunings (when compared to the *master-slave* situation). It was also found that, for moderate coupling and detuning, there was a region between unsynchronized behaviour and oscillation quenching which exhibited imperfect synchronization (shown as black dots in Fig. 7). To the knowledge of the authors this is something that has not been reported before. Apart from this imperfectly locked region, the synchronization regime diagram of our system looks similar to others reported in the literature, such as in [2].

Investigations of mixed states showed (at least for the chosen set of parameters) that moderate coupling strengths stabilized and synchronized the systems (e.g. the periodic system remained in a periodic regime while the chaotic system switched to a periodic regime), whilst with relatively strong coupling the periodic system became chaotic, but the synchronized state was preserved. It would be of interest to investigate whether this behaviour is found more generally in other systems, such as in coupled Rössler and/or Lorenz systems.

4 Laboratory Analogues of Zonally-Symmetric Synchronization

Low order mathematical models of the type discussed in the previous section can provide a great deal of useful insight into the range of possible phenomena that a real physical system might exhibit, provided that physical system can be reasonably assumed to behave as if it was a finite-dimensional dynamical system. For a spatially-extended system, however, such as in a fluid or other continuum, this assumption of restricted dimensionality is by no means guaranteed. For these kinds of system (which, of course, include geophysical systems such as the Earth's atmospheric and oceanic circulation), it is important, therefore, to investigate the applicability of insights obtained from low-dimensional systems with some care.

One approach to this question is to conduct carefully-controlled experiments in the laboratory on fluid flows that encapsulate at least some of the basic physical attributes of spatially-extended geophysical systems. Such experiments are not as easy to implement and control as simplified numerical models, but at least allow the possibility of testing some of the broader implications of simple model studies in a more physical context. Given the difficulties raised above concerning the ambiguity of simple comparisons between model simulations of synchronized behaviour and observed responses in geophysical systems, the capability of laboratory systems to conduct *repeatable* experiments to test particular model hypotheses can be a particular strength.

4.1 Periodic Perturbations

In the present context of baroclinic zonal index cycles, experimental work on externally-perturbed synchronization has recently been performed [34] using a modified version of the so-called thermally driven, rotating baroclinic annulus [23, 24].

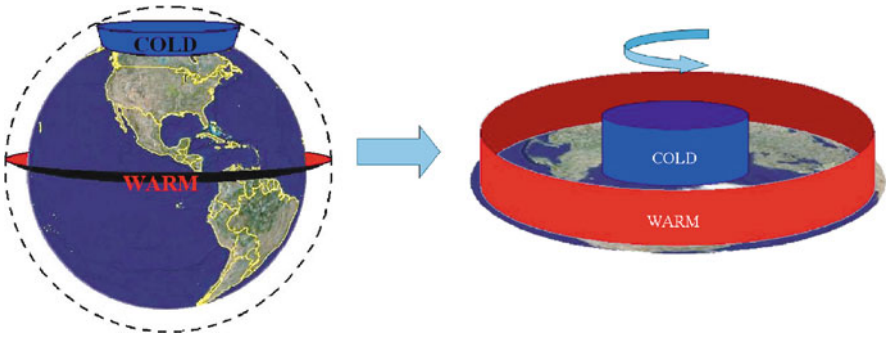


Fig. 8 Atmospheric equivalent of the baroclinic annulus experiment

As shown schematically in Fig. 8, this system consists of a cylindrical annulus of a working fluid confined between two coaxial, thermally-conducting cylinders that are rotated at uniform angular velocity Ω about their common (vertical) axis of symmetry. The combined effect of differential heating (ΔT) in the horizontal direction and uniform background rotation leads to the formation of a zonally-symmetric baroclinic jet flow around the annulus that may become unstable to travelling baroclinic waves under suitable conditions. The system has been extensively studied over many years and is well known to exhibit a wide range of flow regimes under steady rotation and applied thermal contrast ΔT, including pure, axisymmetric flow, steady amplitude travelling waves, periodic amplitude modulated waves and a range of more complex spatiotemporal flows, culminating in forms of geostrophic turbulence. Under weakly unstable conditions, however, the simple low-dimensional mathematical system discussed in the previous section appears to provide a reasonable qualitative model for the first few bifurcations from axisymmetric flow in the experiment.

Like the simple model study discussed in the previous section, the investigation of [34] was focused on the study of the frequency entrainment of the natural amplitude vacillation of the baroclinic wave by an external, zonally-symmetric, cyclic forcing produced by periodically varying the temperature contrast between the inner and outer sidewalls of the annulus. When the forcing was applied to quasi-periodic amplitude-modulated wave regimes, phase locking and frequency entrainment between the applied forcing and the natural frequency of the amplitude modulation was found in a form that closely resembles what was seen in the simple, low-order model of [25]. The frequency range over which such locked states were found varied with the amplitude of the applied forcing, producing a classical Arnol'd tongue (see Fig. 9).

When the detuning was increased, sporadic phase slips and eventually a quasi-periodic behavior was found. Synchronization was even found when forcing was applied with twice the period of the natural modulation frequency. This is illustrated in Fig. 10, which shows the onset of phase-locking with the response occurring at twice the frequency of the imposed forcing. It was therefore concluded

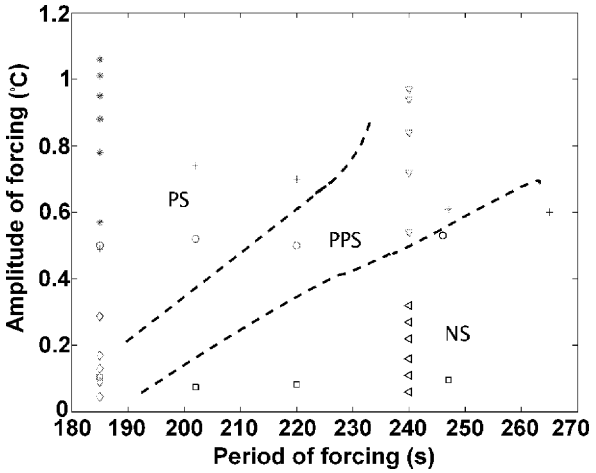


Fig. 9 Parametric variation of partial and complete phase synchronization between the periodic forcing and temperature amplitude of $m = 3$ as a function of amplitude ϵ and frequency ω of the forcing, close to the 1:1 ratio between forcing and the natural vacillation frequencies in the experiment of [34]. *Dotted lines* represent boundaries between different synchronized states: PS denotes full phase synchronization, PPS partial phase synchronization, and NS unsynchronized behavior. Different *symbols* denote different experimental runs

that synchronization of the natural amplitude modulation frequency by zonally-symmetric external forcing is readily achieved, even in the presence of experimental noise and fluctuations in experimental control, suggesting that such synchronized states are robust and potentially observable in nature.

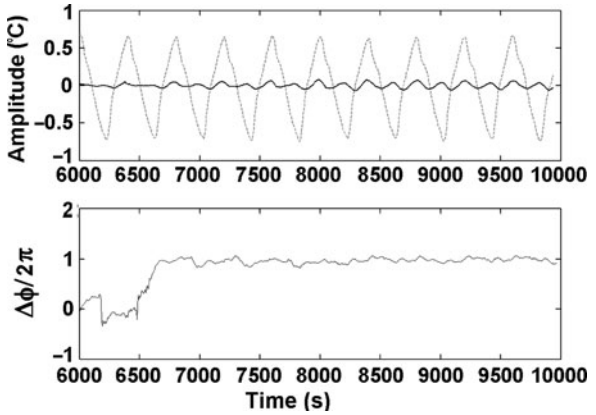


Fig. 10 (**a**) Extract of the temperature variation of the baroclinic wave (*continuous line*) and the imposed external forcing (*broken line*). (**b**) Time variation of the phase difference $\Delta\phi$ between the external forcing and the amplitude modulation of $m = 3$, for periodic forcing with $\epsilon = 0.6$ K and period $\tau = 400$ s. Natural vacillation period $\tau_v = 180$ s

4.2 Mutual Synchronization Experiments

If we regard the thermally-driven rotating annulus system as a well established analogue of mid-latitude cyclogenesis in large-scale atmospheric and oceanic flow, the externally forced experiment discussed above can be considered, to some extent, to emulate an effect that seasonal forcing may have on the dynamics of large-scale planetary waves in the atmosphere. A pair of thermally coupled annulus experiments could then be interpreted as analogous to a "complete" planet, where both hemispheres or two sectors of the same hemisphere (represented by each annulus, cf. Fig. 11) are solely interacting through the equatorial/tropical regions, in the same spirit as Duane et al. (1999).

At the time of writing, this work is still ongoing, although a number of preliminary experiments so far, where either the coupling strength or the de-tuning were varied, have showed clear signs of phase synchronization. In the same way as in the numerical work presented above, synchronization was studied in both periodic and chaotic regimes as well as in mixed states. By varying the mismatch between the systems (by setting slightly different temperature contrasts in each system) we could move from completely uncorrelated behaviour between the two sub-systems, to imperfectly phase-synchronized states (phase slips) and finally to fully synchronized regimes (though the latter was found only in the periodic case). An example of these experimental results showing phase synchronization, visible here via a Lissajous plot, is shown in Fig. 12. The parameters for this particular experiment are, in terms of the thermal Rossby number (Θ) and Taylor number (Ta), $\Theta_M = 0.656$, $\Theta_S = 0.643$, and $Ta_M = Ta_S = 3.778 \times 10^6$, where

$$\Theta = \frac{g\alpha \Delta T D}{\Omega^2 L^2}, \tag{3}$$

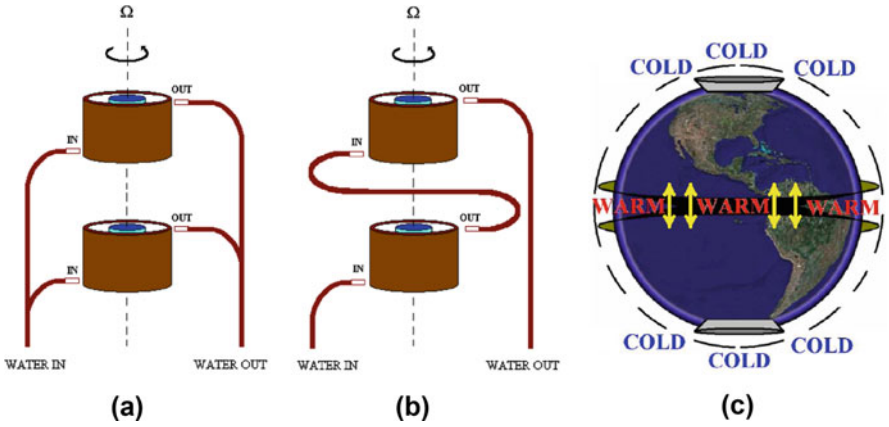


Fig. 11 *Left*: Schematic diagram of the two coupled annuli via the circulating water of the outer circuit. (**a**) Uncoupled, (**b**) Strongly coupled, and (**c**) its geophysical equivalent. *Arrows* represent interhemispheric coupling

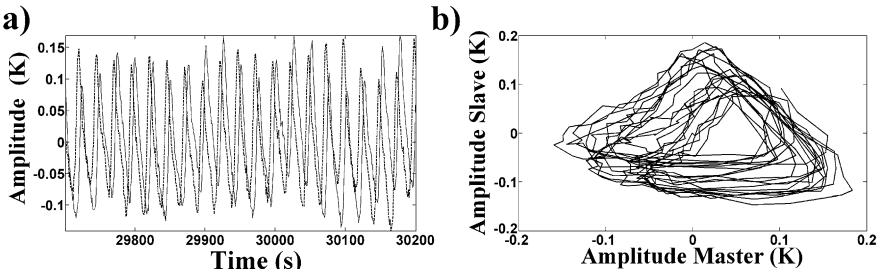


Fig. 12 Two temperature signals originating from the coupled system are shown in (**a**) and its corresponding Lissajous figure in (**b**) (*master* system is shown in a *solid line* and *slave* system in a *dashed line*). Introduction of the coupling between the systems adjusts the frequencies and phases, although the amplitudes remain different. A circular structure in the coupled case in the Lissajous plot is typical of two signals at the same frequency and a constant phase shift

g is the acceleration due to gravity, α the thermal expansion coefficient, D the depth of the annulus, L the gap width and

$$Ta = \frac{4\Omega^2 L^5}{\nu^2 D}. \tag{4}$$

The subindexes "M" and "S" stand for *master* or *slave* systems respectively.

Although the coupling used so far in our experiments is quite schematic when compared with explicit geophysical analogues such as in a baroclinic, midlatitude storm zone in the atmosphere or oceans, there is no reason to believe that the occurrence of synchronization effects is particularly sensitive to the precise way in which the "slave" system is perturbed by the "master", so long as the influence is coherent and mainly via the zonal mean flow (or at least on a larger scale than that of the propagating waves).

The key result, however, is that a clear route to synchronization is found in a real system of fully-developed baroclinic waves, even via relatively weak, zonally-symmetric coupling.

5 Discussion

It is now widely recognized that a thorough understanding of the environment and climate of the Earth, and the development of methods to predict its future behaviour and responses to changes in factors such as atmospheric composition and external forcing, requires a holistic consideration of the entire Earth system. Such an approach views the Earth in terms of a hierarchical ensemble of distinct sub-systems (the troposphere, stratosphere, oceans, cryosphere, land surface and biosphere, for example), all mutually interacting via complex coupling and feedback processes and subject to time-varying external forces (such as the diurnal and annual cycles, and other external processes on longer timescales). Even within these subsystems,

such as in the general circulation of the atmosphere itself, distinct circulation systems and sub-processes can be identified, each of which will interact with other components of the general circulation on various space and time scales, leading to complex behaviours which may be very difficult to predict. Such a view of the Earth System underpins many modern approaches to modelling the Earth's present and past climates, and more recently also to evaluating socio-economic and ecological responses to such changes.

In a system as complex as the Earth, interactions between sub-systems are likely themselves to be highly complex, intermittent and nonlinear, presenting enormous challenges to the modelling community to represent accurately and realistically. In this context, a knowledge and understanding of generic interactions between dynamical systems in the presence of nonlinearity is vital to guide the future development of modelling strategies. Until recently, the approach of the climate science community has tended to focus upon methods based on linearised theory to guide the formulation of modelling and theoretical interpretations of observed phenomena. Until very recently (early $\sim$2000 s), for example, linear cross-correlation techniques have formed the main analysis tools in the study of atmospheric teleconnection patterns, although these essentially only compare the amplitude fluctuations in climate records.

However, methodologies deriving from nonlinear systems are beginning to be more widely recognized, amongst which are those pertaining to synchronized networks and coupled systems. References [35, 36], for example, have shown that nonlinear phase analyses, developed from the study of simple, nonlinear synchronized systems, can reveal more subtle connections between systems when applied to detect correlations between "teleconnected" atmospheric phenomena, such as El Niño and the Indian Monsoon, or rainfall patterns over Oxford and Vienna linked to the NAO. They have demonstrated that some subtle entrainments and coherence associated with higher order synchronization are invisible to conventional correlation techniques and that, in general, nonlinear phase analysis produced better quantitative results, in particular in the estimation of the phase lag/difference between two oscillatory phenomena. It is likely, therefore, that there are many other cyclic phenomena in which some subtle but significant degrees of synchronization could be revealed if studied with modern time series tools based on the phase analysis and phase synchronization approaches. The existence of such synchronization effects may have a strong influence on the behaviour of the system and its potential predictability.

The potential implications of synchronized behaviour, either due to external perturbations or associated with mutual interactions of internal, though geographically separated, components of the climate system may be very substantial, both with regard to the potential predictability of the system (even if it behaves chaotically) and to the interpretation of comparisons between model predictions and observations. The latter especially emphasises the need for careful and repeatable experimentation, as well as more straightforward operational forecast verification against ongoing synoptic analyses of observations.

The possibility that synchronization, and the coupling that leads to associated phase-locking, could itself vary with time is another factor that is only just beginning

to receive attention. But two recent studies [35, 37] have indicated that the onset of synchronization between disparate climate signals can be stimulated by discrete events, such as a major volcanic eruption.

In the immediate future, it would now seem timely to begin to concentrate on somewhat less well studied examples of climate cycles, including those with timescales that are more closely analogous to those in our laboratory experiments. One possible candidate, and perhaps closest in concept to the experimental analogue presented in [32, 34], is the application of synchronization diagnostics to intraseasonal cyclic oscillations in the tropics and extra-tropics. Cyclic variations at tropical latitudes in the troposphere are typically dominated by the well known Madden-Julian oscillation (MJO), with a timescale of around 40–50 days [38, 39]. This oscillation is thought to arise from a coupled atmosphere-ocean wave, energised by atmospheric convection, whose distribution is influenced by patterns in tropical sea surface temperatures (SSTs). However, the detailed mechanism for its occurrence is still not well reproduced in most large-scale atmospheric models, and appears to entail the interaction of processes on a wide range of different scales. This is probably one reason why modelling the MJO has been difficult, and some recent attempts using so-called "aquaplanet" models are just starting to be able to reproduce it, [40]. Oscillations on similar timescales (20–60 days) are also evident in mid-latitude observational timeseries, as revealed, for example, in those of atmospheric angular momentum by [41]. However, it remains unclear as to what relation, if any, these mid-latitude cycles have to the MJO, and what are the mechanisms for such interactions.

In Fig. 13 we show a plot of the latitude-frequency dependence of observed atmospheric angular momentum over a 10 year time interval. This is close to what our idealized, zonally-symmetrically coupled experiment aims to emulate: two mid-latitude (interacting) zones where zonally-symmetric atmospheric cyclic phenomena are developing on a 10–100 day timescale, i.e. somewhat longer than the timescale of individual weather events. In this figure, one can observe that the dominant feature at tropical latitudes is the Madden – Julian Oscillation, with an average period of approximately 50 days. It is also possible to identify various other intraseasonal oscillations in the extra-tropics. For example, in the Southern hemisphere (around 40°S), an oscillation with a period of around 50–60 days is clear. Also, both hemispheres have evident oscillations with periods of approximately 43 days. Other oscillations at several other time scales are also recognizable, such as a feature with a period of ~ 20 days in the Northern Hemisphere and another with a period of 30–31 days in both hemispheres. Even though some of these extra-tropical oscillations have periods very similar to the MJO, it is not yet clear whether or not those phenomena are at all connected. Synchronization diagnostics applied to timeseries of MJO and these extra-tropical intraseasonal indices (similar to those used to produce Fig. 13), for example, could reveal the existence and nature of such a relation and any subtle or intermittent coherence more clearly than in previous work.

With regard to the climate system in general, it is also pertinent to ask why evidence for synchronization in the atmosphere and oceans is not more widespread? It is possible that this lack if evidence is due in part to the use of only linear correlation methods in most studies to date. Imperfect and intermittent forms of phase

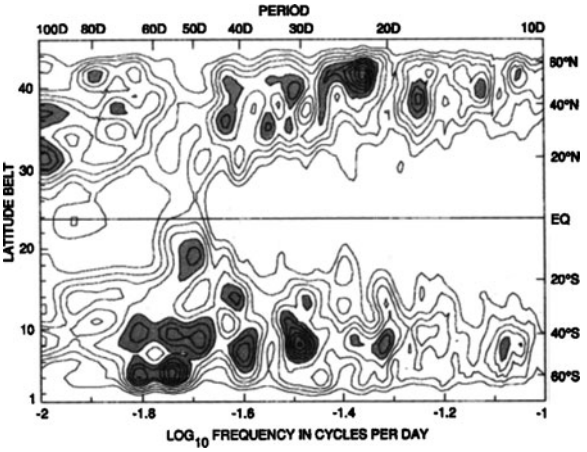


Fig. 13 Latitude-frequency dependence of observed atmospheric angular momentum, as shown by a contour plot of power spectral density. Taken from [41]

synchronization (showing characteristic phase slips), and higher order synchronization states could therefore be underlying the dynamics of some atmospheric phenomena, but have yet to be convincingly detected. There are, for example, roughly quasi-biennial oscillations in both the stratosphere (the well-known QBO, see [8]) and in the upper ocean/troposphere (called the Tropospheric Biennial Oscillation, or TBO, [42]), yet they rarely seem to stay entrained or in phase for very long (see for example [43, 44]). This suggests either that the dynamical coupling between the two processes may be very weak or it may not be constant in time, and therefore, perhaps, active only during part of the year. Such a situation would probably produce frequent phase slips that may be hard to detect as coherent synchronization by traditional methods.

The possible consequences of such an intermittent coupling for the observability of synchronization is another aspect of the problem that could be studied fruitfully within a laboratory experiment such as the coupled rotating annulus system discussed above. Such an extension to the experimental study could allow the quantification of the signature of such a scenario, and hopefully assist in devising a strategy for its detection in climate data.

References

1. A. Pikovsky, M. Rosenblum, and J. Kurths. *Synchronization, A Universal Concept in Nonlinear Sciences*. Cambridge University Press, Cambridge (2001).
2. G. V. Osipov, A. S. Pikovsky, M. G. Rosenblum, and J. Kurths. Phase synchronization effects in a lattice of nonidentical Rössler oscillators. *Phys. Rev. E*, **55**, 2353–2361 (1997).
3. S. Boccaletti, J. Kurths, G. Osipov, D. K. Valladares, and C. S. Zhou. The synchronization of chaotic systems. *Phys. Rep*., **366**, 1–101 (2002).

4. A. E. Hramov, A. A. Koronovskii, and P. V. Popov. Generalized synchronization in coupled Ginzburg-Landau equations and mechanisms of its arising. *Phys. Rev. E*, **72**, 037201, 1–4 (2000).
5. W. F. Ruddiman. *Earth's Climate: Past and Future*. WH Freeman, New York (2001).
6. C. Wunsch. Quantitative estimate of the Milankovitch-forced contribution to observed quaternary climate change. *Quat. Sci. Rev*., **23**, 1001–1012 (2004).
7. J. D. Hays, J. Imbrie, and N. J. Shackleton. Variations in the earth's orbit: pacemaker of the ice ages. *Science*, **194**, 1121–1132 (1976).
8. M. P. Baldwin, L. J. Gray, T. J. Dunkerton, K. Hamilton, P. H. Haynes, W. J. Randel, J. R. Holton, M. J. Alexander, I. Hirota, T. Horinouchi, D. B. A. Jones, J. S. Kinnersley, C. Marquardt, K. Sato, and M. Takahashi. The quasi-biennial oscillation. *Rev. Geophys*., **39**, 179–230 (2001).
9. C. B. Leovy, A. J. Friedson, and G. S. Orton. The quasi-quadrennial oscillation of Jupiter's equatorial stratosphere. *Nature*, **354**, 380–382 (1991).
10. G. S. Orton, P. A. Yanamandra-Fisher, B. M. Fisher, A. J. Friedson, P. D. Parrish, J. F. Nelson, A. S. Bauermeister, L. Fletcher, D. Y. Gezari, F. Varosi, A. T. Tokunaga, J. Caldwell, K. H. Baines, J. L. Hora, M. E. Ressler, T. Fujiyoshi, T. Fuse, H. Hagopian, T. Z. Martin, J. T. Bergstralh, C. Howett, W. F. Hoffmann, L. K. Deutsch, J. E. Van Cleve, E. Noe, J. D. Adams, M. Kassis, and E. Tollestrup. Semi-annual oscillations in Saturn's low-latitude stratospheric temperatures. *Nature*, **453**, 196–199 (2008).
11. W. B. Rossow, A. D. Del Genio, and T. Eichler. Cloud-tracked winds from pioneer venus OCPP images. *J. Atmos. Sci*., **47**, 53–84 (1990).
12. J. Hampson and P. Haynes. Phase alignment of the tropical stratospheric QBO in the annual cycle. *J. Atmos. Sci*., **61**, 2627–2637 (2004).
13. J. Namias. The index cycle and its role in the general circulation. *J. Atmos. Sci*., **7**, 130–139 (1950).
14. W. J. Randel and J. L. Stanford. An observational study of medium-scale wave dynamics in the southern hemisphere summer. Part I:Wave structure and energetics. *J. Atmos. Sci*., **42**, 1172–1188 (1985).
15. D. B. Stephenson, H. Wanner, S. Bronnimann, and J. Luterbacher. The history of scientific research on the North Atlantic Oscillation. *Geophys. Monogr*., **134**, 287–301 (2003).
16. J. M. Wallace. North Atlantic Oscillation/annular mode: two paradigmsone phenomenon. *Quart. J. R. Meteorol. Soc*., **126**, 791–805 (2000).
17. B. J. Hoskins and D. J. Karoly. The steady linear responses of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci*., **38**, 1179–1196 (1981).
18. D. B. Stephenson, V. Pavan, and R. Bojariu. Is the North Atlantic Oscillation a random walk? *Int. J. Clim*., **20**, 1–18 (2000).
19. G. S. Duane, P. J. Webster, and J. B. Weiss. Co-occurrence of northern and southern hemisphere blocks as partially synchronized chaos. *J. Atmos. Sci*., **56**, 4183–4205 (1999).
20. E. Tziperman, M. E. Raymo, P. Huybers, and C. Wunsch. Consequences of pacing the Pleistocene 100 kyr ice ages by nonlinear phase locking to Milankovitch forcing. *Paleoceanography*, **21**, PA4206, 1–11 (2006).
21. G. S. Duane and J. J. Tribbia. Synchronized chaos in geophysical fluid dynamics. *Phys. Rev. Lett*., **86**, 4298–4301 (2001).
22. G. S. Duane. Synchronized chaos in extended systems and meteorological teleconnections. *Phys. Rev. E*, **56**, 6475–6493 (1997).
23. R. Hide and P. J. Mason. Sloping convection in a rotating fluid. *Adv. Phys*., **24**, 47–100 (1975).
24. P. L. Read, M. Collins, W. -G. Früh, S. R. Lewis, and A. F. Lovegrove. Wave interactions and baroclinic chaos: a paradigm for long timescale variability in planetary atmospheres. *Chaos Solitons Fractals*, **9**, 231–249 (1998).
25. F. Eccles, P. L. Read, and T. W. N. Haine. Synchronization and chaos control in a periodically forced quasi-geostrophic two-layer model of baroclinic instability. *Nonlinear Proc. Geophys*., **13**, 23–39 (2006).

26. N. A. Phillips. Energy transformations and meridional cirulations associated with simple baroclinic waves in a two-level quasi-geostrophic model. *Tellus*, **6**, 273–286 (1954).
27. J. Pedlosky. Finite-amplitude baroclinic waves. *J. Atmos. Sci*., **27**, 15–30 (1970).
28. J. Pedlosky. Finite-amplitude baroclinic waves with small dissipation. *J. Atmos. Sci*., **28**, 587–597 (1971).
29. J. Pedlosky. Limit cycles and unstable baroclinc waves. *J. Atmos. Sci*., **29**, 53–63 (1972).
30. A. Lovegrove, I. Moroz, P. L. Read. Bifurcations and instabilities in rotating two-layer fluids: I. f-Plane. *Nonlinear Proc. Geophys*., **8**, 21–36 (2001).
31. A. Lovegrove, I. Moroz, and P. L. Read. Bifurcations and instabilities in rotating two-layer fluids: II. β-plane. *Nonlinear Proc. Geophys*., **9**, 289–309 (2002).
32. A. A. Castrejon-Pita. *Synchronization in baroclinic systems*. DPhil thesis, University of Oxford, Oxford (2008).
33. N. F. Rulkov, M. M. Sushchik, L. S. Tsimring, H. D. I. Abarbanel. Generalized synchronization of chaos in directionally coupled chaotic systems. *Phys. Rev. E*, **51**, 980–994 (1995).
34. F. Eccles, P. L. Read, A. A. Castrejon-Pita, T. W. N. Haine. Synchronization and chaos control of modulated travelling baroclinic waves in a periodically forced, rotating fluid annulus. *Phys. Rev. E*, **79**, 015202(R), 1–4 (2009).
35. D. Maraun and J. Kurths. Epochs of phase coherence between El Niño/Southern Oscillation and Indian Monsoon. *Geophys. Res. Lett*., **32**, L15709, 1–5 (2005).
36. D. Rybski, S. Havlin, and A. Bunde. Phase synchronization in temperature and precipitation records. *Physica A*, **320**, 601–610 (2003).
37. A. A. Tsonis, K. Swanson, and S. Kravtsov. A new dynamical mechanism for major climate shifts. *Geophys. Res. Lett*., **34**, L13705, 1–5 (2007).
38. R. A. Madden and P. R. Julian. Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci*., **28**, 702–708 (1971).
39. R. A. Madden and P. R. Julian. Observations of the 40–50-day tropical oscillation – a review. *Mon. Weather Rev*., **122**, 814–837 (1994).
40. H. Miura, M. Satoh, T. Nasuno, A. T. Noda, and K. Oouchi. A Madden–Julian oscillation event realistically simulated by a global cloud-resolving model. *Science*, **318**, 1763–1765 (2007).
41. J. O. Dickey, M. Ghil, and S. L. Marcus. Extratropical aspects of the 40–50 day oscillation in length-of-day and atmospheric angular momentum. *J. Geophys. Res*., **96**, 22643–22558 (1991).
42. C. -P. Chang and T. Li. A theory for the tropical tropospheric biennial oscillation. *J. Atmos. Sci*., **57**, 2209–2224 (2000).
43. K. F. Kwan and A. A. Samah. A conceptual model relating the quasibiennial oscillation and the tropospheric biennial oscillation. *Int. J. Clim*., **23**, 347–362 (2003).
44. V. Sathiyamoorthy and K. Mohanakumar. Characteristics of tropospheric biennial oscillation and its possible association with the stratospheric QBO. *Geophys. Res. Lett*., **27**, 669–672 (2000).

Stochastic Synchronization

Ram Ramaswamy, R.K. Brojen Singh, Changsong Zhou, and Jürgen Kurths

Abstract The manner in which deterministic nonlinear dynamical systems synchronize differs significantly from the analogous phenomenon in stochastic systems. In this brief review, we discuss the basic issues pertaining to synchrony in the presence of stochasticity. This issue has gained in importance in recent years in part due to the study of finite systems. In particular, biological processes at the cellular and sub cellular level are subject to large fluctuations and nevertheless exhibit significant temporal correlations over a wide range of time-scales. The emergence and maintenance of stochastic synchrony is thus of fundamental importance, and we discuss some illustrative cases.

1 Introduction

Synchronization has been a major theme of studies in nonlinear science since the 1990s. The realization that appropriate coupling can cause deterministic chaotic dynamical systems to completely synchronize [1, 2] catalyzed a number of applications notably those that dealt with communication [3] and control [4]. The major developments in the study of synchronization in nonlinear systems have been summarized in a recent monograph [5] where the different forms of synchrony that are currently realized are discussed in detail. When all the variables of two coupled systems vary identically, they are said to be in *complete* synchrony [5]. The dynamics can be in phase, out of phase, or with a time delay, in which case it is termed *lag* synchrony [6]. In *phase* synchrony [7], the amplitudes are uncorrelated but the phases are locked with respect to each other, and when the variables have a unique functional relation to one another, the systems are said to display *generalized* synchronization [8].

R. Ramaswamy (✉)
Center for Computational Biology and Bioinformatics, and School of Physical Sciences,
Jawaharlal Nehru University, New Delhi 110067, India
e-mail: r.ramaswamy@mail.jnu.ac.in

M. Thiel et al. (eds.), *Nonlinear Dynamics and Chaos: Advances and Perspectives,*
Understanding Complex Systems, DOI 10.1007/978-3-642-04629-2_9,

Recognition of the importance of extending these concepts to stochastic dynamical systems has come from two distinct streams of research. The first deals with synchrony in the presence of (usually experimental) noise, and considerable effort has gone into showing that many of the mechanisms for synchronization are robust to external noise [9]. The second, which is more fundamental, deals with the process of synchronization in systems that are subject to intrinsic noise [10], where the evolution of the systems themselves is contingent on stochastic dynamical processes.

Important examples of this latter class are biological systems which exhibit synchrony at macroscopic as well as microscopic scales [11–13]. Detailed studies of cooperative behaviour at the cellular (or even higher) level, and of the dynamics of subcellular processes in the past several years has established that stochasticity plays an important role in many biological phenomena [14–16]. Similarly, some of the most spectacular instances of synchrony in nature also come from biology: entire populations of fireflies are known to flash synchronously [17] and a population of cells, say yeast, can be made to divide synchronously [18]. The manner in which such synchrony arises can be traced in different cases to different causes: inter-organism communication or quorum sensing [19] being two well known means. The existence of circadian clocks in organisms ranging from cyanobacteria [20] to mammals [21] and the consequent adjustment of internal processes to the day – night cycle suggests that there are a variety of different mechanisms that can be effective in synchronizing stochastic systems.

Extensive studies of deterministic dynamical systems, largely in the past two decades, have shown that in general, synchronization occurs in one of the following scenarios.

- The simplest is when identical dynamical systems are mutually weakly coupled [22], usually linearly. The case of nonlinear coupling has also been considered, more as a consequence of the manner in which specific systems may interact- for instance, the coupling of neurons [23].
- Chaotic synchronization is most commonly associated with one-way coupling, the so-called "master-slave" scenario [2]: one of the systems drives the other and is consequently unaffected by the dynamics of the response (see Fig. 1 for an example).
- A third situation is when two identical systems are subject to an common drive. In this case the two systems exhibit mutual synchronization and their response to the drive is termed generalized synchrony [8].

As has also been extensively studied in the past decades, the systems under consideration need not be identical; the synchronization phenomenon persists under a fair level of parameter mismatch.

For stochastic systems, the above definitions – indeed the basic notions – need to be modified and in the past few years, a number of different studies have addressed this issue, both from the point of view of uncovering the essential strategies by which systems can synchronize, to quantitative and analytical studies on the effect of noise and stochasticity on the dynamics.

The term stochastic synchronization has itself been used quite variably in the literature, although always in reference to dynamical systems where the evolution

has a noise-governed component [24, 25]. This can be either additive or multiplicative noise, and common examples of where these considerations are of importance is in neuronal dynamics or in other biological systems. Almost any experimental situation has some level of noise that is unavoidable. Given this context, therefore, it is also commonly associated with the phenomenon of stochastic resonance, and a number of studies have discussed the parallels between the two phenomena [26].

Two of the central issues that are germane to the notion of synchrony in stochastic systems were already pointed out over 40 years ago by Stratonovich [27]. One is that stochastic systems can properly only show phase synchronization rather than complete synchronization, and the other, that inherent fluctuations can interfere with the maintenance of synchronization: stochastic systems can drift in and out of phase synchrony if the noise is sufficiently intense.

What does it mean for stochastic systems to synchronize? Consider deterministic nonlinear oscillators for which a phase can be defined quite generally. Two such coupled oscillators have the equations of motion

$$\frac{d\mathbf{x}_i}{dt} = F_i(\mathbf{x}, \alpha) + \epsilon G_i(\mathbf{x}_i, \mathbf{x}_j), \quad i, j = 1,2. \tag{1}$$

The phase synchronized state can then be defined rigorously by first defining phases $\phi_i(t)$ for the two (sub)systems, and requiring that they maintain a constant well-defined phase difference as a function of time. Clearly, something analogous can be defined for stochastic dynamics as well so long as there is first a sense of oscillation, and secondly, a phase-like variable. Fortunately, in a number of stochastic systems, these conditions are met.

The present subject is one whose scope is vast, with potential applicability in disciplines ranging from nanoscience [28] to economics [29]. In this review, however, we will confine our attention to the following aspects: (i) What is the effect of noise on synchrony, and (ii) how can stochastic systems be coupled in order that they synchronize.

In the following section, we describe measures for the detection of synchrony in stochastic systems before proceeding to discuss noise-induced synchronization in Sect. 3. Section 4 is devoted to the analysis of different mechanisms for the coupling of stochastic systems in order to cause them to synchronize. Examples from the recent literature that pertain to systems and synthetic biology are reviewed. The article concludes with a summary in Sect. 5.

2 Measures for Stochastic Synchronization

We first address the issue of how the synchronization of stochastic signals may be detected (or judged). In general, quantitative measures for assessing the synchronization of *any* two different signals, say $s_1(t)$ and $s_2(t)$, are commonly based on the distance between them, or on some function of this distance, $d_{s_1,s_2}(t) = ||s_1(t) - s_2(t)||$. In the synchronous state, the signals are identical and $d_{s_1,s_2} \to 0$,

and the trajectory of the coupled system is confined to a lower dimensional subspace, the synchronization manifold [5]. In situations where phase synchronization obtains, one considers the phases of the two signals, namely $\phi_1(t)$ and $\phi_2(t)$, and phase locking is achieved when

$$\Delta\phi = m\phi_1(t) - n\phi_2(t) = \text{Constant}. \tag{2}$$

The above definitions which are inspired by the study of linear or nonlinear oscillators can be extended to the case of chaotic signals as well although there may or may not be a sense of "phase" or indeed of "oscillation". Indeed, as shown by Gabor [30], given any signal $s_j(t)$, it is possible to define its phase through the so-called analytical signal approach. The Hilbert transform of $s_j(t)$ is given by

$$\tilde{s}_j(t) = \frac{1}{\pi}\text{PV}\int_{-\infty}^{+\infty} \frac{s_j(t)}{t-\tau}d\tau \tag{3}$$

where PV denotes the Cauchy principal value. The instantaneous phase $\phi_j(t)$ and amplitude $A_j(t)$ can then be defined through the following identity,

$$s_j(t) + i\tilde{s}_j(t) = A_j(t)\exp i\phi_j(t). \tag{4}$$

When the signals $s_1(t)$ and $s_2(t)$ are the outcomes of stochastic processes, the above definitions need adaptation. The distance between the two signals is unlikely to be zero but is bounded within a limit that depends on the level of stochasticity,

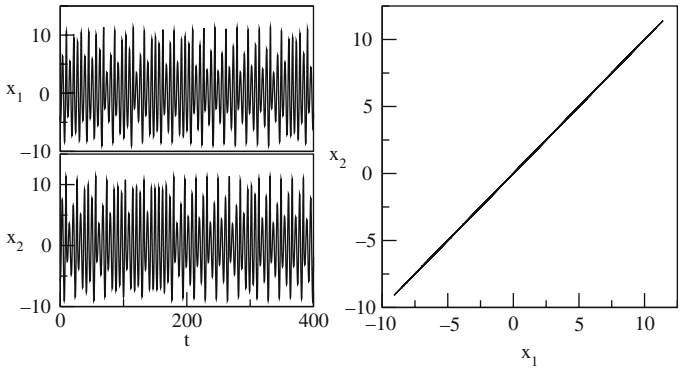


Fig. 1 Chaotic synchronization of two Rössler oscillators within the master-slave coupling scenario. Starting from different initial conditions in the phase space, the trajectories of two systems, shown in (**a**) and (**b**) rapidly synchronize, converging onto the synchronization manifold as shown in (**c**). Standard parameters are taken[1], and the coupling is as indicated in [2]

[1] In standard notation [5] the parameters of the two Rössler oscillators are $a = 0.2$, $b = 0.2$, $c = 5.7$, and the drive variable is y.

and thus the nature of synchronization in the presence of stochasticity is in a sense blurred. It is often clearer to judge phase synchrony since the Hilbert phase can be computed in a straightforward manner, but as already indicated by Stratonovich [27] the phase difference between two signals can also drift owing to intrinsic noise [10].

One may also use the idea of a similarity function, namely

$$S(\tau) = \sqrt{\frac{\langle [s_2(t+\tau) - s_1(t)]^2 \rangle}{\langle s_1^2(t) \rangle \langle s_2^2(t) \rangle}} \tag{5}$$

or its generalizations [6] to examine whether this quantity reaches a sufficiently low minimum for specific τ; this would allow for the detection of lag synchrony between stochastic signals. Some (but not all) studies of stochastic systems have used the above criteria which can clearly provide suitable quantitative measures for detecting synchronization.

3 The Effect of Stochasticity on Synchrony

As discussed above, there are a number of common coupling scenarios through which two systems might synchronize [5]. Further, the nature of the synchrony itself can be complete [2], phase [7], lag [6], or generalized [8]. To judge the robustness of the synchronization, it is customary to subject the coupled systems to external perturbations such as additive noise [5]. This is crucial if the synchrony is to be detected in experiments, and its importance has been appreciated in most of these studies. Within a master-slave setting, noise or drift in the drive can induce similar drift in the response, although synchronization in this scenario is very robust, persisting till high levels of noise [31].

There is another source of stochasticity in this context. Practically speaking, two systems can never be identical, and thus when considering an ensemble of coupled systems, this dispersion in system parameters should also be taken into account. It has been shown that in general, the deviation from synchrony depends on both the degree to which the parameters are not identical, the strength of the noise, and its correlation properties.

In general, in a population of N oscillators,

$$\dot{\mathbf{X}}_\mathbf{i} = \mathbf{F}(\mathbf{X}_\mathbf{i}) + \epsilon \mathbf{G}(\mathbf{X}) + \sigma \xi_i(t), \;\; i = 1, \ldots, N \tag{6}$$

will synchronize for appropriate coupling $\mathbf{G}$, for sufficiently large ϵ. When noise ξ of amplitude σ is added to each unit of an assembly of synchronized oscillators the usual effect is to cause desynchronization [32, 33] if the noise is uncorrelated, namely if $\langle \xi_i(t) \xi_j(t) \rangle = 0$ for $i \neq j$. The behaviour is different, depending on what the dynamics of the noise-free system is: when this is periodic, the effect is less pronounced than when the motion is chaotic.

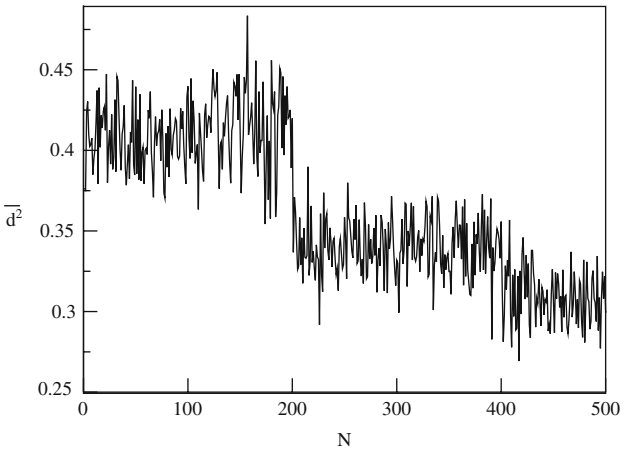


Fig. 2 Convergence of chaotic orbits in the logistic mapping $x \to 4x(1-x) + \eta$ under identical additive noise η. The distance of two trajectories subject to the same uniform random noise distributed in the interval $[-W,W]$ is computed, subjecting these to successively increasing noise (see [34] for details). The average distance between pairs of trajectories decreases with time, displaying increased order under the influence of additive noise

A surprising result is that when uncoupled systems ($\epsilon = 0$) are subject to *identical noise*, $\xi_i \equiv \xi$, this leads, somewhat counterintuitively, to increased ordering. The operative caveat is that the uncoupled systems should be identical as well, but starting with studies of mappings and flows that were subject to stochastic modulation [34], a considerable body of work has been built up to show that this effect is robust. An instance of synchronization by common noise of chaotic orbits in the logistic map is shown in Fig. 2 (after [34]). As consequence of noise-induced order is that uncoupled systems that are subject to *identical* noise will respond in the same ordered state, namely they will synchronize, and a number of studies have addressed the mechanisms underlying noise-induced synchrony.

This effect, of noise-induced complete synchronization of identical chaotic systems, has led to some controversy. The conclusion that strong enough noise can synchronize chaotic systems appears oxymoronic, and explanations ranged from finite precision effects to the role of bias, namely the nonzero mean of applied noise. However, as shown by Zhou and Kurths [35], the mechanism leading to complete synchronization by a common additive noise is the existence of a region in the phase space where there is significant contraction. Further, if the nature of stochastic modulation is such that a given system can occasionally explore those regions of phase space that give rise to negative local Lyapunov exponents, namely converging dynamics, then noise will be able to induce statistical phase synchronization.

Shown in Fig. 3 are the Lyapunov exponents as a function of noise intensity in model chaotic oscillators subject to common noise [35]. In the case of Rössler oscillators, there is no significat contraction region, so noise induced synchrony cannot be realized here (Fig. 3a, c), unlike the Lorenz case (Fig. 3b, d). Noise-induced synchronization is very likely to be observed in homoclinic chaotic systems: in such

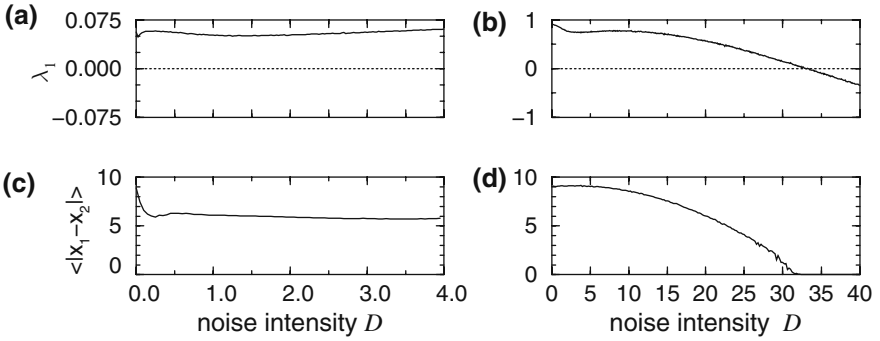


Fig. 3 The largest Lyapunov exponents λ_1 and average synchronization error between two oscillators with common noise as a function of the noise intensity D. (**a**) and (**c**) In the case of two Rössler oscillators, there is no noise-induced synchrony, while in (**b**) and (**d**) two Lorenz systems synchronize for $D \geq 31$. From [35]

systems, chaotic trajectories pass by an unstable fixed point (a saddle), leaving the fixed point along the unstable manifold, and reinject along a strongly contracting stable manifold. By adding low amplitude noise, synchronization of two such identical systems can easily be realized since the orbit spends a smaller fraction of the time along the unstable manifold. As a result, noise not only induces synchronization between the oscillators, but can also enhances temporal coherence of the dynamics [36]. This is shown in Fig. 4 for a model laser system for which both theory and experiment are possible. Noise-enhanced coherence can improve the response and synchronization to a periodic external drive [37]. The nontrival effects of added noise have been demonstrated in a chaotic spiking neuronal model [38] in addition to the homoclinic chaotic CO_2 laser [36, 37].

In weakly coupled chaotic dynamical systems (namely when the coupling alone is itself not strong enough to induce significant synchronization) external noise may enhance synchronization. For example, in two coupled nonidentical chaotic oscillators, the phases of the oscillators can be correlated, but this phase-locking is disrupted when the chaotic trajectories approach unstable periodic orbits that have very different periods. In this case, a suitable amount of noise may *prevent* the systems approaching such "unlocking" orbits. In this manner, phase slips can be avoided, effectively enhancing synchronization (see Fig. 5). Such a noise-induced effect has been shown in both numerical simulations of the Rössler oscillators as well as experiments with electrochemical oscillators [39].

In an ensemble of weakly coupled oscillators, correlated noise can generate significant collective oscillations that are not available just from the coupling alone [40, 41]. Coherence of the collective oscillation is maximal at an intermediate noise level (as in the case of stochastic resonance; see below) depending on the coupling parameters. Similar mechanisms have been shown to operate in causing synchronization in spatially extended systems [42], and indeed noise enhanced collective synchronization could have meaningful applications, for instance in harvesting energy from turbulent oceanic waves using coupled oscillatory electrical generators.

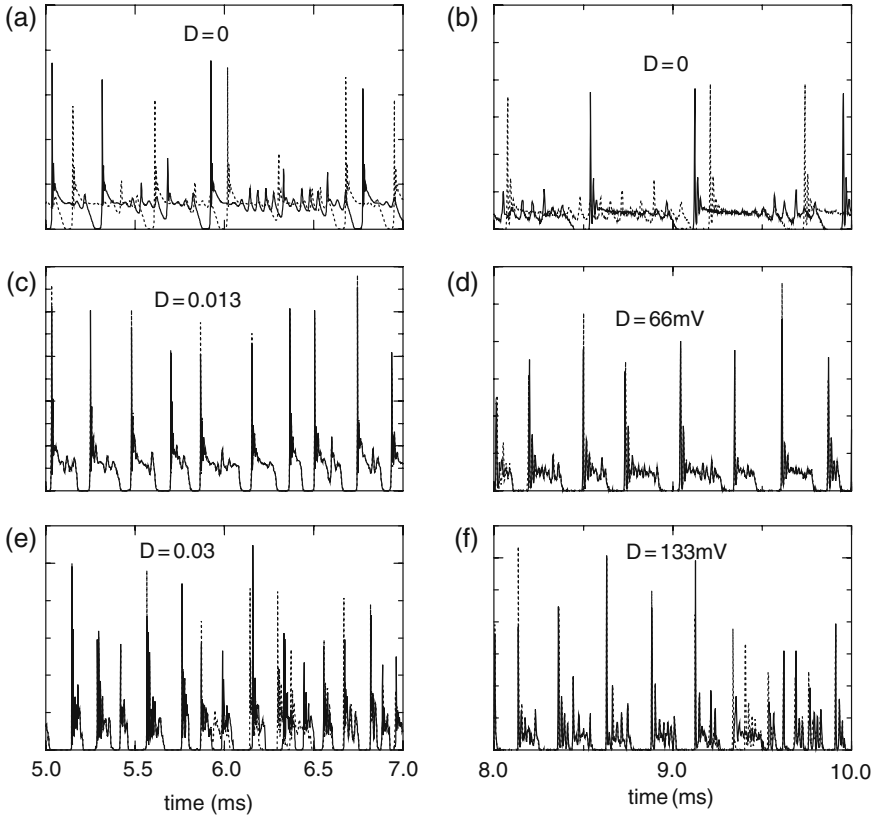


Fig. 4 Time series of output intensities (arbitrary units) of two laser models x_1 (*solid lines*) and y_1 (*dotted lines*) with identical parameters, a common noise at various intensities and an independent noise (intensity $D_1 = 0.0005 \sim$ intrinsic noise level). Note that that the spiking interval becomes shorter and more regular, and at the same time, the two lasers obtain synchronization in the presence of a sufficient amount of common noise. *Left panel*: model system, *right panel*: experiment. From [37]

An alternate viewpoint considers noise-induced synchrony as a manifestation of generalized synchronization between the oscillators and the noise [43]. Recall that a response system, say with variables $\mathbf{Y}$, is termed in generalized synchrony with a drive (with variables $\mathbf{X}$) when the dynamics of the response is a unique function of the drive, namely

$$\mathbf{Y} = \Phi(\mathbf{X}). \tag{7}$$

It is not clear how far this holds when the drive itself is a stochastic variable, especially since the phenomenon persists when there is both common as well as independent noise, as was demonstrated recently in a study of an ensemble of uncoupled oscillators subject to both forms of noise [44]. The dynamics of each oscillator is given by

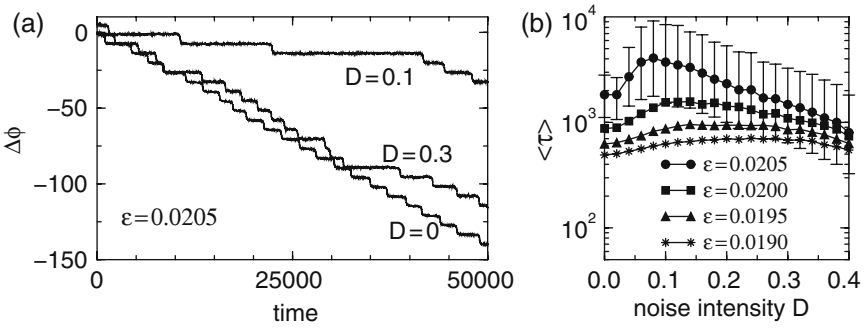


Fig. 5 Noise-enhanced phase locking in two weakly coupled Rössler chaotic oscillators. (**a**) Phase difference vs. time for different noise intensity D. (**b**) Average duration of phase-locking epochs vs. D for different coupling strength ϵ. The standard deviation is shown with error bars for coupling $\epsilon = 0.0205$. From [38]

$$\dot{\mathbf{X}}_\mathbf{i} = \mathbf{F}(\mathbf{X_i}) + \epsilon \mathbf{G}(\mathbf{X_i})\xi(t) + \sigma \mathbf{H}(\mathbf{X_i})\eta_\mathbf{i}(t), \tag{8}$$

and it was shown that irrespective of details of the dynamics, for sufficiently weak noise, synchronization and phase coherence always results.

Noise-induced synchronization has features of similarity with the general phenomenon of stochastic resonance (SR) [45], the amplification of weak signals due to noise. SR, first described by Benzi, Sutera and Vulpiani [46] has been invoked in a number of studies of systems subject to noise. A linear response theory is adequate in its analysis, for the case of weak noise, and when the dynamics of the system can be modeled by two states, with noise induced switching between them, then SR has also been described as a stochastic synchronization of the switching with the noise-indeed, as a phase *coherence* that is increased as a consequence of noise [45].

4 The Emergence of Synchrony in Stochastic Systems

When considering the dynamics of microscopic systems, the nature of fluctuations may be large enough to preclude a purely deterministic description of the phenomena. Examples can be drawn from a number of fields, in particular from biology: cellular or subcellular processes typically involve a small number of molecules interacting in a finite volume. Indeed, a stochastic formulation of chemical kinetics has been developed since the 1950s [47]. A detailed consideration of fluctuations is germane when dealing with confined geometries of any kind- biological cells or nanoscale reactors. On this scale, the variation of dynamical quantities is governed by stochastic evolution, and the system is properly described through the master equation formalism [48].

In the typical case one may consider a set of variables $X_1, X_2, \ldots, X_N$ that undergo M possible "reactions".

$$X_i + X_j + \cdots \xrightarrow{k_\mu} X_\alpha + X_\beta + \ldots, \quad \mu = 1, \ldots, M. \tag{9}$$

Depending on the process at hand, the variables X_i can correspond to actual molecular species, or can be defined more generally. The k_μ's are the rates of the different channels, and the instantaneous configuration of the system, $\mathcal{C}$ is defined by specifying the values of the different variables, namely $X_1, X_2, \ldots, X_N$.

Denoting the configurational probability at time t by $P(\mathcal{C};t)$, the evolution of the system is governed by the master equation

$$\frac{dP(\mathcal{C},t)}{dt} = -\sum_{\mathcal{C}'} P(\mathcal{C},t) W_{\mathcal{C}\to\mathcal{C}'} + \sum_{\mathcal{C}} P(\mathcal{C}',t) W_{\mathcal{C}'\to\mathcal{C}} \tag{10}$$

where the W's are transition probabilities depending on the stochastic processes that are possible [48]. Solving the master equation gives a specific trajectory in the phase space of the problem, but since the system is stochastic, any particular realization starting from some $\mathcal{C}_0$ will only give a probabilistic description of the system as it travels through the phase space. Although solving the master equation is typically very difficult, for specific systems both analysis and Monte Carlo simulations can be carried out [49, 50]. In the limit of large volumes and large numbers of the different species, it can be shown rigorously that the above master equation reduces to the usual mass-action kinetics laws, with additional noise terms that inversely depend on the system volume. This gives the so-called chemical Langevin equation [51], a set of coupled stochastic differential equations of the general form

$$\frac{dX_i(t)}{dt} = \sum_{j=1}^{M} s_{ji} a_j(\mathbf{X}(t)) + \sum_{j=1}^{M} s_{ji} a_j(\mathbf{X}((t))^{1/2} \gamma_j(t) \tag{11}$$

where the $\gamma_j(t)$'s are statistically independent uncorrelated Gaussian white noise while the s_{ij} are stoichiometric factors particular to the reaction channels at hand, with a_j's being related to the "reaction" propensities [49]. When the noise can be ignored completely, this reduces to the usual mass-action kinetic equations,

$$\frac{dX_i(t)}{dt} = \sum_{j=1}^{M} s_{ji} a_j(\mathbf{X}(t)), i = 1, \ldots, N. \tag{12}$$

However, it is in the limit of small systems that stochasticity plays a major role, and thus the dynamical evolution via Eq. (10) is relevant. Given two such similar *noninteracting* stochastic systems, their dynamics will be totally uncorrelated. However, a number of coupling strategies that bring about stochastic synchrony between the two have been discussed in the recent literature [52]. For convenience if the variables of the two systems are indicated with and without primes, then the configurations are $\mathcal{C} \equiv (X_1, X_2, \cdots, X_N)$ and $\mathcal{C}' \equiv \left(X_1', X_2', \cdots, X_N'\right)$ respectively.

1. When *directly* coupled the two systems have a variable in common, say $X_i \equiv X_i'$. Then the dynamics of the remaining variables X_j and X_j' (for all j) become essentially identical. It can be shown that in the limit of large volumes when the master equation reduces to the corresponding mass-action kinetic laws, namely a set of deterministic coupled differential equations, this becomes similar to the master-slave coupling discussed by Pecora and Caroll [2].
2. In *exchange* coupling, the two systems are linked by the additional stochastic processes,

$$X_i \xrightarrow{c} X_i' \tag{13}$$

$$X_i' \xrightarrow{c'} X_i \tag{14}$$

   where c and c' are the respective transition rates. Synchronization in the other variables, namely between X_j and X_j' (for all j) occurs when the rates c and c' are sufficiently large. In the large volume limit, this essentially corresponds to bidirectional diffusive coupling.
3. *Global* or *mean-field* coupling is another effective strategy to couple two or more similar systems through a common variable. This provides a "mean-field" through which the different systems communicate, causing the remaining variables to exhibit synchrony [53, 54].

While these are not the only means through which stochastic systems can synchronize, a number of recent studies have discussed applications that exploit one or the other of the above. In the remainder of this section we discuss some representative cases.

A simple example is afforded by the coupling of two chemical oscillators [52]. The Brusselator [55] is the set of reactions

$$\begin{aligned} A_1 &\xrightarrow{c_1} X \\ A_2 + X &\xrightarrow{c_2} Y + A_3 \\ 2X + Y &\xrightarrow{c_3} 3X \\ X &\xrightarrow{c_4} A_4. \end{aligned} \tag{15}$$

and for appropriate choices of the elementary rates, the c_i's, the dynamics of the variables X and Y in the above system is on a noisy limit-cycle [49].

Consider two such Brusselators (with primed and unprimed variables distinguishing them). In the case of direct coupling $X \equiv X'$, namely the species X is common to both subsystems. With exchange coupling, as indicated above there is an additional channel,

$$X \xrightarrow{c} X' \tag{16}$$

$$X' \xrightarrow{c'} X \tag{17}$$

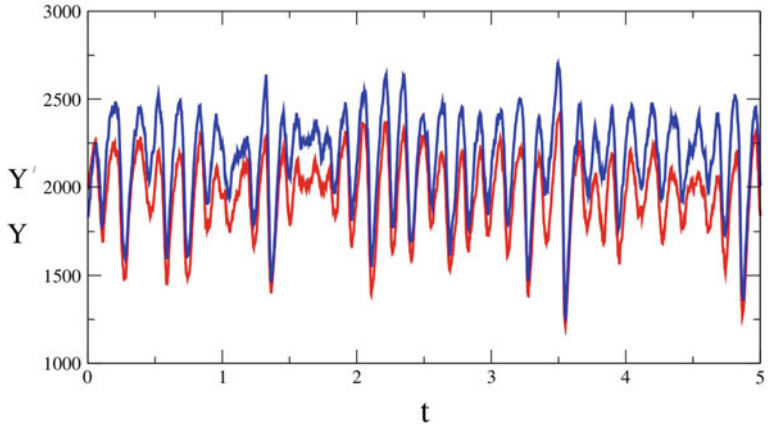


Fig. 6 Phase synchrony in directly coupled Brusselators. There is 10% mismatch in the reaction constants $c_2 = 50$ and $c_2' = c_2 + 5$, and the other parameters are $c_1 = c_1' = 16 \times 5{,}000$, $c_3 = c_3' = 0.00005$ and $c_4 = c_4' = 16 \times 5$. Adapted from [52]

through which the subsystems communicate. As has been shown [52] for coupled Brusselators or for more complex genetic oscillators (that are appropriate models for actual circadian oscillators [56, 57]) with either form of the coupling scheme, the variables Y and Y' will phase synchronise even for a high level of parameter mismatch. Shown in Fig. 6 is a representative case of phase synchrony: although the volumes of the coupled systems are quite different and the fluctuations are large, the variables of the two systems are highly correlated in their variation although the amplitudes clearly differ considerably.

Exchange coupling is effective only above a threshold exchange rate, namely for sufficiently large c and c' in Eq. (16) and (17). Thus if one examines the phase difference, $\Delta\phi$ between the two coupled oscillators, this should become zero only above a critical value of the coupling. Shown in Fig. 7 are results for the above system, showing that for low coupling the phase difference between the systems drifts, becoming zero only for large c. Indeed, such an effect had been noted in an earlier simulation of such systems using the chemical Langevin equations [58, 51].

Mean-field coupling has been extensively examined in a related setting [59]. A biological phenomenon that essentially uses such a mean-field is quorum sensing, namely the (concentration-dependent) activation of specific genetic switches by signaling molecules that are themselves produced by other genes. Early studies that suggested that such forms of coupling could give rise to synchrony included a study of globally coupled repressilators [60], a model regulatory network with a quorum-sensing module [61], as well as a more abstract circadian model pertaining to the multinuclear organism *Neurospora crassa* [54]. In this model, one of the species is the cytosolic clock gene mRNA denoted M and a clock protein denoted P_C. The nuclear proteins are denoted $P_N^{(i)}$ the superscript indicating which cell it is from. The dynamics of the individual nuclear genetic oscillators are synchronized as a consequence of terms of the form

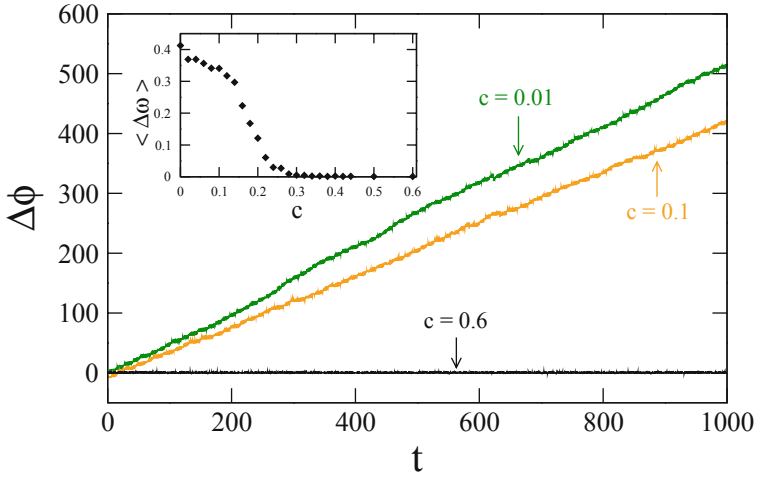


Fig. 7 The drift of the phase difference between two exchange-coupled Brusselators. The inset shows the variation of the slope as a function of the coupling strength, namely the exchange reaction rate $c = c'$. Adapted from [52]

$$\frac{dP_C}{dt} \sim F(M, P_C) + k \sum_{i=1}^{N} P_N^{(i)} \tag{18}$$

in the evolution equations, where N is the number of nuclei, F contains the functional dependence particular to this model, and k is the coupling of the mean field provided by the nuclear proteins [54]. A more specific model of such coupling is via the process of quorum sensing, effected in a scheme (see Fig. 8) elaborated by Zhou, Chen and Aihara [61] who showed that extracellular noise common to all cells can effectively induce collective dynamics and stochastic synchronization.

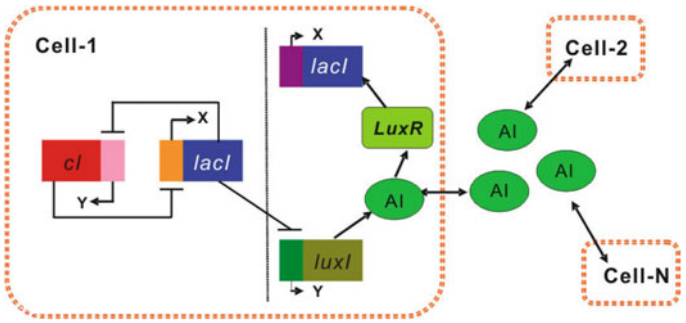


Fig. 8 Cellular communication is provided by the diffusing *AI* dimers in the extracellular environment; *AI* is synthesized by the protein *LuxI*, production of which is inhibited by the $AI_2 - LuxR_2$ tetramer which is itself formed from the *AI* dimer and the *LuxR* dimer. (Courtesy of the authors [61])

5 Discussion and Summary

In this chapter we have addressed two aspects of stochastic synchronization. The first is to do with the inducement of synchrony by stochasticity, namely the manner in which systems can exploit the ordering effects of noise so as to achieve synchronous dynamics. This is of relevance in a variety of biological settings where the synchronization of systems can arise through their being subject to a common noise or other external perturbations [62]. As has often been pointed out, in ecology, the environment can play a great role in bringing synchrony to different populations over large geographical regions. Experiments have been carried out, particularly in neuronal systems, and show that this is particularly important for information processing since a group of neurons can respond collectively to common synaptic events.

A second focus has been on the synchronization of systems with intrinsic noise. These studies are again of relevance to biological systems and indeed to any microscopic finite system. Recent applications have been to the study of synchrony in genetic oscillators and other networks, and a particularly interesting recent experiment uses these ideas to synchronize chemical oscillators in a nanoscale environments [63].

Synchrony is one the most visible effects of coupling between nonlinear dynamical systems [22]. It is also ubiquitous [5], and is a very general means through which temporal organization arises in complex systems [64]. Its persistence—or enhancement—in the presence of both internal and external noise is thus of importance both from fundamental as well as applications points of view.

Acknowledgement It is a pleasure to dedicate this article to Celso Grebogi who has been a major influence in nonlinear science for the past several decades, and a major inspiration as well.

References

1. H. Fujisaka and T. Yamada. Stability theory of synchronized motion in coupled-oscillator systems. *Prog. Theor. Phys*., **69**, 32 (1983).
2. L. M. Pecora and T. L. Caroll. Synchronization in chaotic systems. *Phys. Rev. Lett*., **64**, 821 (1990).
3. K. M. Cuomo and A. V. Oppenheim. Circuit implementation of synchronized chaos with applications to communications. *Phys. Rev. Lett*., **71**, 65 (1993); G. D. Van Wiggeren and R. Roy. *Science*, **279**, 1198 (1998).
4. J. M. Gonz'alez-Miranda. *Synchronization and Control of CHAOS*. World Scientific, Singapore (2004).
5. A. S. Pikovsky, M. G. Rosenblum, and J. Kurths. *Synchronization: A Universal Concept in Nonlinear Science*. Cambridge University Press, Cambridge (2001).
6. M. G. Rosenblum, A. S. Pikovsky, and J. Kurths. From phase to lag synchronization in coupled chaotic oscillators. *Phys. Rev. Lett*., **78**, 4193 (1997).
7. M. G. Rosenblum, A. S. Pikovsky, and J. Kurths. Phase synchronization of chaotic oscillators. *Phys. Rev. Lett*., **76**, 1804 (1996).
8. N. F. Rulkov, M. M. Sushchik, L. S. Tsimring, and H. D. I. Abarbanel. Generalized synchronization of chaos in directionally coupled chaotic systems. *Phys. Rev. E*, **51**, 980

(1995); V. S. Afraimovich, N. N. Verichev, and M. I. Rabinovich. Stochastic synchronization of oscillation in dissipative systems. *Radiophys. Quantum Electron*., **29**, 795(1986).
9. L. Kocarev, U. Parlitz, and R. Brown. Robust synchronization of chaotic systems. *Phys. Rev. E*, **61**, 3716 (2000).
10. N. G. Van Kampen. *Stochastic Process in Physics and Chemistry*. North-Holland, Amsterdam (1992).
11. A. T. Winfree. *The Geometry of Biological Time*. Springer Verlag, New York (2001).
12. L. Glass. Synchronization and rhythmic processes in physiology. *Nature*, **410**, 277 (2001); L. Glass and M. C. Mackey. *From Clocks to Chaos. The Rhythms of Life*. Princeton University Press, Princeton (1988).
13. A. Goldbeter. Computational approach to cellular rhythms. *Nature*, **420**, 238 (2002).
14. L. J. S. Allen. *An Introduction to Stochastic Processes with Applications to Biology*. Prentice-Hall, NJ (2003).
15. P. S. Swain, M. B. Elowitz, and E. D. Siggia. Intrinsicand extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. (USA)*, **99**, 12795 (2002).
16. H. H. McAdams and A. Arkin. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet*., **15**, 65 (1999).
17. J. Buck. Synchronous rhythmic flashing of fireflies. II. *Quart. Rev. Biol*., **63**, 265 (1988) for a review.
18. J. Aldridge and E. K. Pye, Cell density dependence of oscillatory metabolism. *Nature* **259**, 670 (1976). This is now a standardized experimental protocol, as discussed in D. C. Amberg, D. J. Burke, and J. N. Strathern. *Methods in Yeast Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (2005). See also A. De Monte, F. d'Ovidio, S. Dano, and P. G. Sorensen. Dynamical quorum sensing: Population density encoded in cellular dynamics. *Proc. Natl. Acad. Sci. (USA)*, **104**, 18377 (2007).
19. A. Camilli and B. Basler. Bacterial small-molecule signaling path ways. *Science*, **311**, 1113 (2006).
20. T. Kondo, T. Mori, N. V. Lebedeva, S. Aoki, M. Ishiura, and S. S. Golden. Circadian rhythms in rapidly dividing cyanobacteria. *Science*, **275**, 224 (1997).
21. S. M. Reppert and D. R. Weaver. Coordination of circadian timing in mammals. *Nature*, **418**, 935 (2002).
22. C. Huygens. *Horoloqium Oscilatorium*. Parisiis, Paris (1673).
23. A. Prasad, M. Dhamala, B. M. Adhikari, and R. Ramaswamy, Amplitude death in nonlinear oscillators with nonlinear coupling. *Phys. Rev.* E (2010, in press).
24. V. S. Afraimovich, N. N. Verichev, and M. I. Rabinovich. Stochastic synchronization of oscillation in dissipative system. *Radioph. Quant. Elec*., **29**, 1573–9210 (1986).
25. A. Neiman, A. Silchenko, V. Anishchenko, and L. Schimansky-Geier. Stochastic resonance: Noise-enhanced phase coherence. *Phys. Rev. E*, **58**, 7118 (1998).
26. A. Neiman, L. Schimansky-Geier, A. Cornell-Bell, and F. Moss. Noise-enhanced phase synchronization in excitable media. *Phys. Rev. Lett*., **83**, 4896 (1999); A. Neiman, L. Schimansky-Geier, F. Moss, B. Shulgin, and J. J. Collins. Synchronization of chaotic systems by stochastic signals. *Phys. Rev. E*, **60**, 284 (1999).
27. R. L. Stratonovich. *Topics in the Theory of Random Noise*. Gordon and Breach, New York (1967).
28. R. M. Nieminen and A. E. J. Jansen. Monte Carlo simulations of surface reactions *Appl. Catal. A*, **160**, 99 (1997).
29. J.-P. Loy and C. R. Weiss. Synchronisation due to common shocks? Evidence from German Grocery Prices *Econ. Lett*., **85**, 123 (2004).
30. D. Gabor. Theory of communication. *J. IEE Lond*., **93**, 429 (1946).
31. R. Brown, N. F. Rulkov, and N. B. Tuffilaro. Synchronization of chaotic systems: The effects and additive noise and drift in the dynamics of the driving. *Phys. Rev. E*, **50**, 4488 (1994).
32. D. S. Goldobin and A. S. Pikovsky. Synchronization and desynchronization of self-sustained oscillators by common noise. *Phys. Rev. E*, **71**, 045201R (2005).

33. J. Teramae and D. Tanaka. Robustness of the noise-induced phase synchronization in a general class of limit cycle oscillators. *Phys. Rev. Lett*., **93**, 204103 (2004).
34. A. Maritan and J. Banavar. Chaos, noise, and synchronization. *Phys. Rev. Lett*., **72**, 1451 (1994).
35. C. S. Zhou and J. Kurths. Noise-induced phase synchronization and synchronization transition in chaotic oscillators. *Phys. Rev. Lett*., **88**, 230602 (2002).
36. C. S. Zhou, J. Kurths, E. Allaria, S. Boccaletti, R. Meucci, and F. T. Arecchi. Constructive effects of noise in homoclinic chaotic system. *Phys. Rev. E*, **67**, 066220 (2003).
37. C. S. Zhou, J. Kurths, E. Allaria, S. Boccaletti, R. Meucci, and F. T. Arecchi. Noise-enhanced synchronization of homoclinic chaos in a CO_2 laser. *Phys. Rev. E*, **67**, 015205R (2003).
38. C. S. Zhou and J. Kurths. Noise-induced synchronization and coherence of a Hodgkin-Huxley model of thermally sensitive neurons. *Chaos*, **13**, 401 (2003).
39. C. S. Zhou, J. Kurths, I. Z. Kiss, and J. L. Hudson. Noise-enhanced phase synchronization of chaotic oscillators. *Phys. Rev. Lett*., **89**, 014101 (2002).
40. I. Z. Kiss, Y. Zhai, J. L. Hudson, C. S. Zhou, and J. Kurths. Noise-enhanced phase synchronization and coherence resonance in sets of chaotic oscillators with weak global coupling. *Chaos*, **13**, 267 (2003).
41. C. S. Zhou and J. Kurths. Spatiotemporal coherence resonance of phase synchronization in weakely coupled chaotic oscillators. *Phys. Rev. E*, **65**, 040101R (2002).
42. L. Baroni, R. Livi, and A. Torcini. Transition to stochastic synchronization in spatially extended systems. *Phys. Rev. E*, **63**, 036226 (2001).
43. S. Guan, Y.-C. Lai, C.-H. Lai, and X. Gong. Understanding synchronization induced by "common noise". *Phys. Letts. A*, **353**, 30 (2005).
44. H. Nakao, K. Arai, and Y. Kawamura. Noise-induced synchronization and clustering in ensembles of uncoupled limit-cycle oscillators. *Phys. Rev. Lett*., **98**, 284202 (2007).
45. L. Gammaitoni, P. Hanggi, P. Jung, and F. Marchesoni. Stochastic resonance. *Rev. Mod. Phys*., **70**, 223 (1998).
46. R. Benzi, A. Sutera, and A. Vulpiana. The mechanism of stochastic resonance. *J. Phys*., **A 14**, L453 (1981).
47. D. A. McQuarrie. Stochastic approach to chemical kinetics *J. Appl. Probl*., **4**, 413 (1967).
48. I. Oppenheim, K. E. Shuler, and G. H. Weiss. Stochastic and deterministic formulation of chemical rate equations. *J. Chem. Phys*., **50**, 460 (1969).
49. D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem*., **81**, 2340 (1977).
50. D. T. Gillespie. A general method for numerically simulation the Stochastic time evolution of coupled chemical reactions. *J. Comp. Phys*., **22**, 403 (1976).
51. D. T. Gillespie. The chemical Langevin equation. *J. Chem. Phys*., **113**, 297 (2000); D. T. Gillespie. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem*., **58**, 35 (2007).
52. A. Nandi, G. Santhosh, R. K. B. Singh, and R. Ramaswamy. Effective mechanisms for synchronization of stochastic oscillators. *Phys. Rev. E*, **76**, 041136 (2007); A. Nandi and R. Ramaswamy. Synchronization of coupled Stochastic oscillators: The effect of topology. *Pramana J. Phys*., **70**, 1065 (2008); R. K. Brojen Singh and R. Ramaswamy, to be published.
53. M. G. Rosenblum and A. S. Pikovsky. Controlling synchronization in an ensemble of globally coupled oscillators. *Phys. Rev. Lett*., **92**, 114102 (2004).
54. D. Gonze and A. Goldbeter. Circadian rhythms and molecular noise. *Chaos*, **16**, 026110 (2006).
55. I. Prigogine and R. Lefever. Symmetry breaking instabilities in dissipative systems *J. Chem. Phys*., **48**, 1655 (1968).
56. D. Gonze, J. Halloy, and A. Goldbeter. Robustness of circadian rhythms with respect to molecular noise. *Proc. Natl. Acad. Sci. (USA)*, **99**, 673 (2002).
57. J. M. Vilar, H. Y. Kueh, N. Barkai, and S. Leibler. Mechanism of noise-resistance in genetic oscillators. *Proc. Natl. Acad. Sci. (USA)*, **99**, 5988 (2002); N. Barkai and S. Leibler. Circadian clocks limited by noise. *Nature*, **403**, 267 (2000).

58. M. Wang, Z. Hou, and H. Xiu. Internal noise-enhanced phase synchronization of coupled chemical chaotic oscillators *J. Phys. A*, **38**, 145 (2005).
59. L. Chen, R. Wang, T. Zhou, and K. Aihara. Noise-induced coperative behavior in a multicell system. *Bioinformatics*, **21**, 2722 (2005); C. Li, L. Chen, and K. Aihara. Transient resetting: A novel mechanism for synchrony and its biological examples. *PLoS Comp. Biol*., **2**, e103 (2006); C. Li, L. Chen, and K. Aihara. Stochastic synchronization of genetic oscillator networks. *BMC Syst. Biol*., **1**, 6 (2007).
60. J. Garcia-Ojalvo, M. B. Elowitz, and S. H. Strogatz. Modeling a synthetic multicellular clock: Repressilators coupled by quorum sensing. *Proc. Natl. Acad. Sci. (USA)*, **101**, 10955 (2004).
61. T. Zhou, L. Chen, and K. Aihara. Molecular communication through Stochastic synchronization induced by extracellular fluctuations. *Phys. Rev. Lett*.. **95**, 178103 (2005); T. Zhou, J. Zhang, Z. Yuan, and L. Chen. Synchronization of genetic oscillators. *Chaos*, **18**, 037126 (2008).
62. F. Varela, J.-P. Lachaux, E. Rodriguez, and J. Martinerie. The brainweb: Phase synchronization and large-scale integration. *Nat. Neurosci*., **2**, 229 (2001).
63. A. F. Taylor, M. R. Tinsley, F. Wang, Z. Huang, and K. Showalter. Dynamical Quorum sensing and synchronization in large population of chemical oscillators. *Science*, **323**, 614 (2009).
64. H. Haken. *Synergetics: An Introduction*. Springer Verlag, Berlin (1983).

Experimental Huygens Synchronization of Oscillators

Alexander Pogromsky, David Rijlaarsdam, and Henk Nijmeijer

Abstract We present an experimental set-up that allows to study both controlled and uncontrolled synchronization between a variety of different type of oscillators. The setup consists of two fully actuated mass-damper-spring oscillators mounted on an actuated platform. By means of different types of feedback realized through computer controlled actuation it is possible to demonstrate different synchronization phenomena, i.e. synchronization of pendula, synchronization of Duffing oscillators, synchronization of rotating bodies, etc. Two experiments are presented where uncontrolled synchronization between two types of identical oscillators is investigated. First, uncontrolled synchronization between two Duffing oscillators is investigated and second, uncontrolled synchronization between two coupled rotating disks is discussed.

1 Introduction

In the 17th century the Dutch scientist Christiaan Huygens observed a peculiar phenomenon when two pendula clocks, mounted on a common frame, seemed to "sympathize" as he described it [1]. What he observed was that both clocks adjusted their rhythm towards anti-phase synchronized motion. This effect is now known as frequency or Huygens synchronization and is caused by weak interaction between the clocks due to small displacements of the connecting frame. In [2–5] an extended analysis of this phenomenon is presented. In [6] the authors present an experimental study of Huygens synchronization and finally, in [7, 8] a study of the uncontrolled as well as the controlled Huygens experiment is presented.

Many more cases of synchronization have been identified in nature and technology around us [9, 10]. Two striking examples in biology are the synchronized

A. Pogromsky (✉)
Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands
e-mail: A.Pogromsky@tue.nl

Dedicated to Celso Grebogi at the occasion of his 60th birthday

M. Thiel et al. (eds.), *Nonlinear Dynamics and Chaos: Advances and Perspectives,* Understanding Complex Systems, DOI 10.1007/978-3-642-04629-2_10,

flashing of fireflies [11] or synchronization of neurons in the brain when performing perceptual tasks. Other examples of synchronization can be found in [12, 13]. Using synchronization it is possible to stabilize the frequency of a high power generators and there are more other applications we are unable to mention in this paper.

Three centuries after Huygens the phenomenon of synchronizing driven pendula is, to our best knowledge, repeated twice experimentally. In the first research by Bennett, Schatz, Rockwood and Wiesenfeld [2], one has tried to accurately reproduce the findings of Huygens in an experimental setup consisting of two pendulum clocks attached to a free moving cart. The results of this experiment confirm the documented observations of Huygens. A rather simple but interesting experiment is described by Pantaleone [3], where the synchronization of two metronomes is discussed, which are coupled by a wooden board rolling on soda cans. The metronomes in this setup would synchronize most of the time with in phase oscillations.

In this paper we present an experimental set-up [4] that allows to study both controlled and uncontrolled synchronization between a variety of different oscillators. In Sects. 2 and 3 we outline the goal of the experimental set-up based on our previous experiments. In Sect. 4 the set-up is introduced and we present the dynamical properties of the system. Furthermore, we present the means by which we are able to modify these properties to represent a variety of different oscillators. Next, in Sect. 5, we present an experiment of the synchronization of two Duffing oscillators. We analyze the stability of the synchronization manifold and continue with numerical and experimental results. Section 6 presents an experiment where the set-up is adjusted to behave like two rotating eccentric discs which are coupled through a third disc mounted on a common axis. Conclusions and future research are presented in Sect. 7.

2 Synchronization of Pendulum Clocks

Before giving the description of a new experimental set-up we briefly outline experimental studies performed on a simple set-up that consists of two metronomes located on a common beam. Analysis of those experiments motivated us to develop a new set-up that will be described in the subsequent sections. The experimental setup consists of two metronomes coupled by a platform which can translate horizontally. The metronomes are made by Wittner, type Maelzel (series 845). The platform is suspended by leaf springs, which allows a frictionless horizontal translation. A photograph of the experimental setup is given in Fig. 1.

Several experiments are performed in order to gain experience with the dynamics of the system. Parameters which can be varied in the experiment are the mass of the platform, the mass and frequency of the metronomes and the amount of damping in the system.

With different parameters the experiments show different phenomena: anti-phase synchronization (see Fig. 2), an intermediate (neither anti- nor in-phase

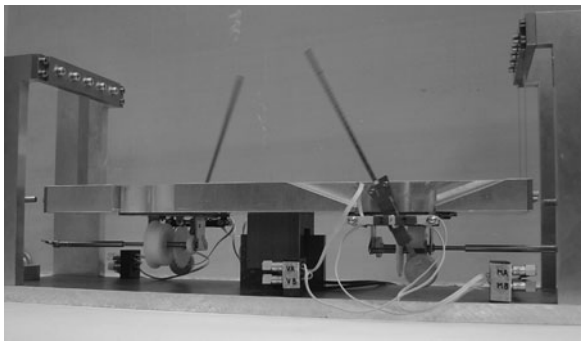


Fig. 1 Photograph of the setup

synchronization) regime with a large amplitude difference of the angles of the metronomes and finally in-phase synchronization (see Fig. 3), for details, see [6].

Though it was possible to demonstrate different synchronization regimes with such a relatively simple set-up, the experiments revealed difficulties that are unavoidable for this set-up. First of all, the experiments showed that the results depend heavily on the escapement mechanism used in the metronomes. This mechanism forces pendulums to oscillate and behave like pendulum clocks. Those mechanisms are relatively difficult to model and their parameters vary for different metronomes causing non-identical synchronization. To change the damping parameter the leaf springs should be replaced and hence the whole set-up should be mechanically rebuilt that resulted in that some experiments are quite difficult to repeat with the same results. Moreover, the set-up allows to perform experiments with only one type of oscillators—pendulums, while Huygens' synchronization can be observed for different types of oscillators, not necessarily pendulums. Analyzing these problems it was proposed to design and to build a new synchronization set-up that would be free of those disadvantages.

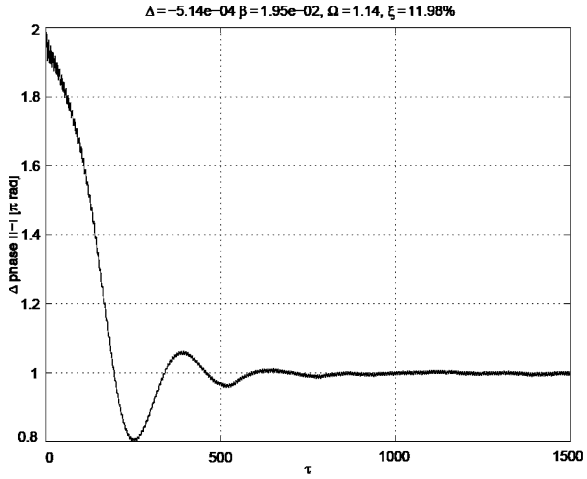


Fig. 2 Anti-phase synchronization. Difference of phases of the oscillators versus time

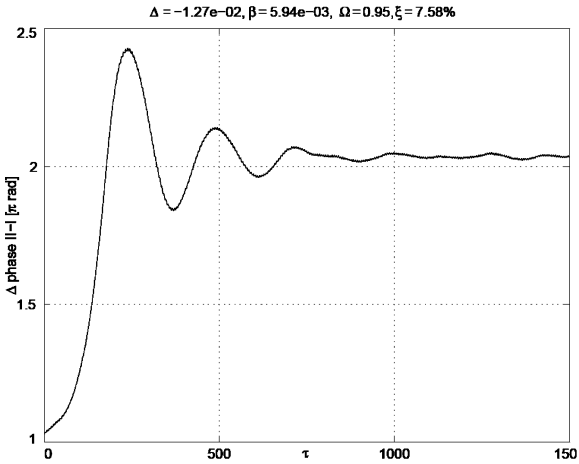


Fig. 3 In-phase synchronization. Difference of phases of the oscillators versus time

3 The Goal of the Experimental Set-Up

Starting from the work of Huygens synchronization phenomena attracted attention of many researchers due to its numerous applications, see e.g. [9, 12, 13]. It turns out that the synchronous behavior can be observed in many practical situations and it demonstrates a certain degree of robustness, that is certain approximate synchronization can happen even if the synchronizing oscillators are not identical. Moreover, understanding of this phenomenon is far from being complete. For example, Christian Huygens observed synchronous behavior of pendulum clocks attached to a common support beam—this effect can be explain by modeling the beam as a one-degree of freedom solid body as is the case in the simple set-up of the previous section, while, to our best knowledge, no attempt was made to analyze this situation taking into account the beam flexibility, reported by Huygens. To discover new aspects of the synchronization phenomenon it would be useful not only to focus on mathematical models and/or computer simulation but also to perform practical experiments to validate the theoretical findings. Moreover, the experiments can also indicate new and interesting directions for theoretical analysis. Since the synchronization can be observed in a variety of different systems it is interesting to have an experimental set-up that can be used to observe and analyze the synchronous behavior for rather different systems. This set-up is useful for both educational and research purposes.

To meet this objective it was decided to design an electro-mechanical set-up in such a way that its mechanical counterpart is capable to model common features of the synchronizing systems – two one-degree of freedom oscillators mounted on a one-degree of freedom common support body (beam, or platform), while by actuating the oscillators and the beam via computer-controlled feedback it would be possible to mimic a variety of different controlled/uncontrolled synchronizing

systems. To this end, the mechanical part was designed as linear as possible, while all the nonlinearities necessary to observe the synchronization are implemented by the electrical part of the set-up. Such a design allows to perform numerous experiments on different systems with limited implementation costs.

The main objectives of the experimental studies are

- To observe experimentally synchronous behavior in a variety of different mechanical systems
- To analyze robustness of the synchronization in a real environment
- To find common features of the synchronizing systems on the uniform experimental platform

4 The Experimental Set-Up

In order to experimentally study synchronization between coupled oscillators a set-up consisting of two oscillators, mounted on a common frame has been developed (see Figs. 4 and 5). The parameters of primary interest are presented in Table 1. The set-up contains three actuators and position sensors covering all degrees of freedom.

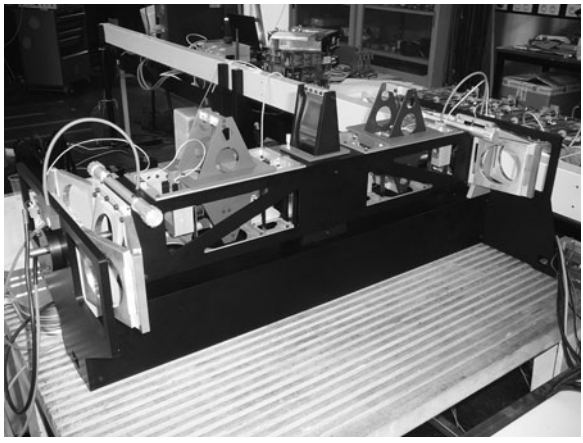


Fig. 4 Photograph of the set-up

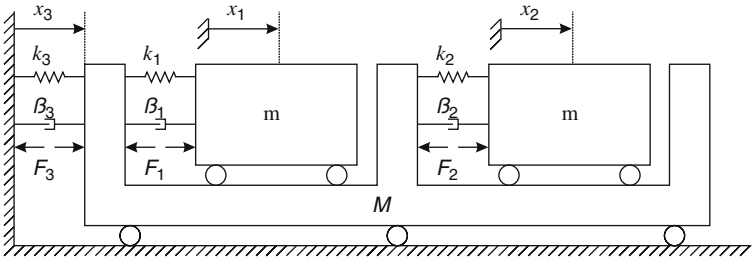


Fig. 5 Schematic representation of the set-up

Table 1 Parameters in experimental set-up

	Oscillator 1	Oscillator 2	Frame/beam
Mass	m	m	M
Stiffness	$\kappa_1(\cdot)$	$\kappa_2(\cdot)$	$\kappa_3(\cdot)$
Damping	$\beta_1(\cdot)$	$\beta_2(\cdot)$	$\beta_3(\cdot)$

Furthermore, although the masses of the oscillators (m) are fixed, the mass of the connecting beam (M) may be varied by a factor 10. This allows for mechanical adjustment of the coupling strength. A schematic representation of the set-up is depicted in Fig. 5 and the equations of motion of the (ideal) system are:

$$m\ddot{x}_1 = -\kappa_1(x_1 - x_3) - \beta_1(\dot{x}_1 - \dot{x}_3) + F_1 \tag{1}$$

$$m\ddot{x}_2 = -\kappa_2(x_2 - x_3) - \beta_2(\dot{x}_2 - \dot{x}_3) + F_2 \tag{2}$$

$$\begin{aligned} M\ddot{x}_3 = {} & \kappa_1(x_1 - x_3) + \kappa_2(x_2 - x_3) \\ & -\kappa_3(x_3) + \beta_1(\dot{x}_1 - \dot{x}_3) + \beta_2(\dot{x}_2 - \dot{x}_3) - \beta_3(\dot{x}_3) \\ & +F_3 - F_1 - F_2, \end{aligned} \tag{3}$$

where m, $M \in \mathbb{R}_{>0}$ and $x_i \in \mathbb{R}$, $i = 1,2,3$ are the masses and displacements of the oscillators and the beam respectively. The functions κ_i:$\mathbb{R} \mapsto \mathbb{R}$, $\beta_i : \mathbb{R} \mapsto \mathbb{R}$ describe the stiffness and damping characteristics present in the system. F_i are the electric actuator forces that may be determined such that the experimental set-up models a large variety of different dynamical systems (see Sect. 4.1).

The stiffness and damping in the real system are found to be very well approximated by:

$$\kappa_i(q_i) = \sum_{j=1}^{5} k_{ij} q^j \tag{4}$$

$$\beta_i(\dot{q}_i) = b_i \dot{q}_i, \tag{5}$$

where $q_1 = x_1 - x_3$, $q_2 = x_2 - x_3$ and $q_3 = x_3$. The values of k_{ij} and b_i $\forall$ $i = 1,2,3$ have been experimentally obtained and will be used to modify the systems' properties in the sequel.

4.1 Adjustment of the System Properties

In order to experiment with different types of oscillators, the derived properties (stiffness and damping) are adjusted. Note that, since we know the damping and stiffness present in the system and since we can fully measure the state of the system, we may adjust these properties, using actuators, to represent any dynamics we want. This allows modeling of different types of springs (linear, cubic) and any other desired effect within the limits of the hardware. In the next part of this paper we present two examples of this type of modulation. The system is first adapted

to analyze synchronization between Duffing oscillators and secondly to analyze the synchronizing dynamics of two coupled rotating eccentric discs under the influence of gravity.

5 Example 1: Coupled Duffing Oscillators

In this section experimental results with respect to two synchronizing Duffing oscillators are presented. After introducing the dynamical system analysis of the limiting behaviour of the system is presented. Finally, both numerical and experimental results are presented and discussed.

5.1 Problem Statement and Analysis

Consider the system as depicted in Fig. 6, where

$$\frac{\kappa_d(q_i)}{m} = \omega_0^2 q_i + \vartheta q_i^3 \tag{6}$$

where $q_i = x_i - x_3$ and constants $\omega_0, \vartheta \in \mathbb{R}_{>0}$.

The system under consideration represents two undriven, undamped Duffing oscillators coupled through a third common mass. The set-up depicted in Fig. 5 can be adjusted to model this system by defining the actuator forces as:

$$F_i = \kappa_i(q_i) + \beta_i(\dot{q}_i) - \kappa_d(q_i), i = 1,2 \tag{7}$$

$$F_3 = 0 \tag{8}$$

Where $F_3 = 0$ is chosen because, in the original set-up, the beam already models the situation as depicted in Fig. 6 (linear stiffness and damping) fairly accurately. The equations of motion of the resulting system are linear:

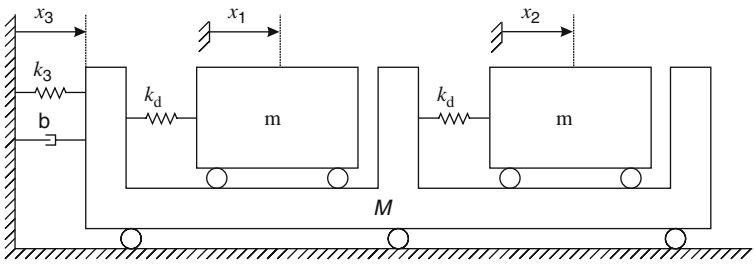


Fig. 6 Schematic representation of the set-up modeling two coupled Duffing oscillators

$$m\ddot{x}_1 = -\kappa_d(x_1 - x_3) \tag{9}$$

$$m\ddot{x}_2 = -\kappa_d(x_2 - x_3) \tag{10}$$

$$M\ddot{x}_3 = \kappa_d(x_1 - x_3) + \kappa_d(x_2 - x_3) - kx_3 - b\dot{x}_3, \tag{11}$$

where k,$b \in \mathbb{R}_{>0}$ are the stiffness and damping coefficients of the beam.

Before continuing with the experimental and numerical results the system's limiting behaviour is analyzed. In order to do so the notion of anti-phase synchronization needs to be defined. We call the solutions of $x_1(t)$ and $x_2(t)$ asymptotically synchronized in *anti-phase* if they satisfy the following ralation

$$\lim_{t\to\infty} ||x_1(t) - \alpha\sigma_{(\frac{T}{2})}x_2(t)|| = 0, \tag{12}$$

with α a scale factor and $\sigma_{(\frac{T}{2})}$ a shift operator over half an oscillation period (T). Respectively, if instead of the previous relation it follows that

$$\limsup_{t\to\infty} ||x_1(t) - \alpha\sigma_{(\frac{T}{2})}x_2(t)|| \leq \varepsilon, \tag{13}$$

for some small $\varepsilon > 0$, we say that the solutions are *approximately* asymptotically synchronized in anti-phase.

It can been shown that the dynamics of the oscillators in (9), (10) and (11) converges to anti-phase synchronization as $t \to \infty$.

To prove this claim consider the system (9), (10) and (11). To analyze the limit behaviour of this system, the total energy is proposed as a candidate Lyapunov function:

$$\mathcal{V} = \frac{1}{2}\sum_{i=1}^{3} m_i\dot{x}_i^2 + \sum_{i=1}^{3}\int_0^{\xi_i} \kappa_i(s)\, ds, \tag{14}$$

where $m_1 = m_2 = m$, $m_3 = M$, $\xi_i = x_i - x_3$, $i = 1,2$, $\xi_3 = x_3$, $\kappa_i(q_i) = \kappa_d(q_i)$ and $\kappa_3 = kx_3$. Calculating the time derivative of $\mathcal{V}$ along the solutions of the system (9), (10) and (11) yields:

$$\dot{\mathcal{V}} = -b\dot{x}_3^2. \tag{15}$$

Hence, we find $\dot{\mathcal{V}} \leq 0$ and the system may be analyzed using LaSalle's invariance principle.

Equation (15) implies that $\mathcal{V}$ is a bounded function of time. Moreover, $x_i(t)$ is a bounded function of time and will converge to a limit set where $\dot{\mathcal{V}} = 0$. On this limit set $\dot{x}_3 = \ddot{x}_3 = 0$, according to (15). Substituting this in system (9), (10) and (11) yields $x_3 = 0$ on the system limit set. Substituting $x_3 = \dot{x}_3 = \ddot{x}_3 = 0$ in (11) shows:

$$\kappa_d(x_1) = -\kappa_d(x_2) \tag{16}$$

Since κ_d is a one-to-one, odd function, this implies:

$$x_1 = -x_2 \tag{17}$$

Finally, substituting $x_1 = -x_2$ in (9) and (10) yields:

$$\dot{x}_2 = -\dot{x}_1. \tag{18}$$

Summarizing, it has been shown that any solution of (9), (10) and (11) will converge to anti-phase synchronized motion according to definition the given definition.

The next paragraph will present numerical and experimental results that support the analysis provided in this section.

5.2 Experimental and Numerical Results

In order to experimentally investigate the synchronizing behaviour of two coupled Duffing oscillators the set-up has been modified as specified in the previous section. The oscillators are released from an initial displacement of -3 and -2.5 mm respectively (approximately in phase) and allowed to oscillate freely.

Figure 7 shows the sum of the positions of the oscillators and the position of the beam versus time. As becomes clear from Fig. 7, approximate anti-phase synchronization occurs within 40 s. Furthermore, Fig. 8 shows the limiting behaviour of both oscillators and the beam. Although the amplitudes of the oscillators differ significantly, the steady state phase difference is 1.01π. The most probable cause for the amplitude difference is the fact that the oscillators are not exactly identical. As a result, the beam does not come to a complete standstill, although it oscillates with an amplitude that is roughly ten times smaller than that of the oscillators.

In addition to the experimental results, numerical results are provided in Figs. 9 and 10. The parameters in the simulation are chosen as shown in Table 2. The results presented in Figs. 9 and 10 correspond to the experimental results provided in Figs. 7 and 8 respectively. Although the oscillation frequencies of the oscillators are almost equal (within 5%) in the simulation and the experiment, the final amplitudes of the oscillators differs by a factor 15. This is due to the fact that in the experiment the damping is over-compensated, resulting in larger amplitudes of the oscillators. In the numerical simulation almost exact anti-phase synchronization with equal oscillator amplitudes is achieved.

Finally, note that some of the differences between the experimental and simulation results may be coped with by tuning either the parameters of the numerical simulation or those of the set-up itself. *The question of identifying a model can thus be reversed to tuning the parameters of the set-up rather than those of the model.*

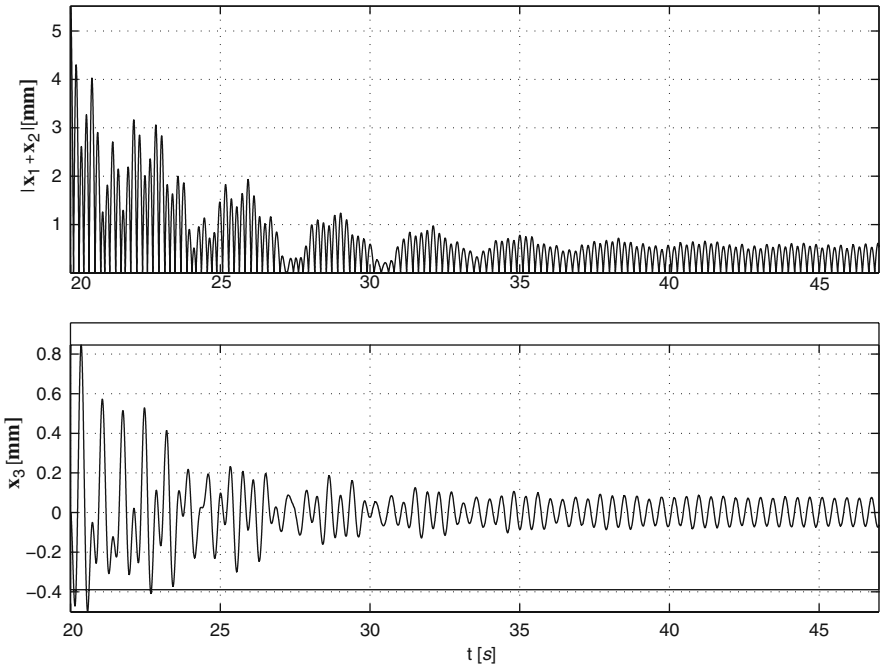


Fig. 7 Experimental results: (*top*) Sum of the displacements of both oscillators. (*bottom*) Displacement of the connecting beam

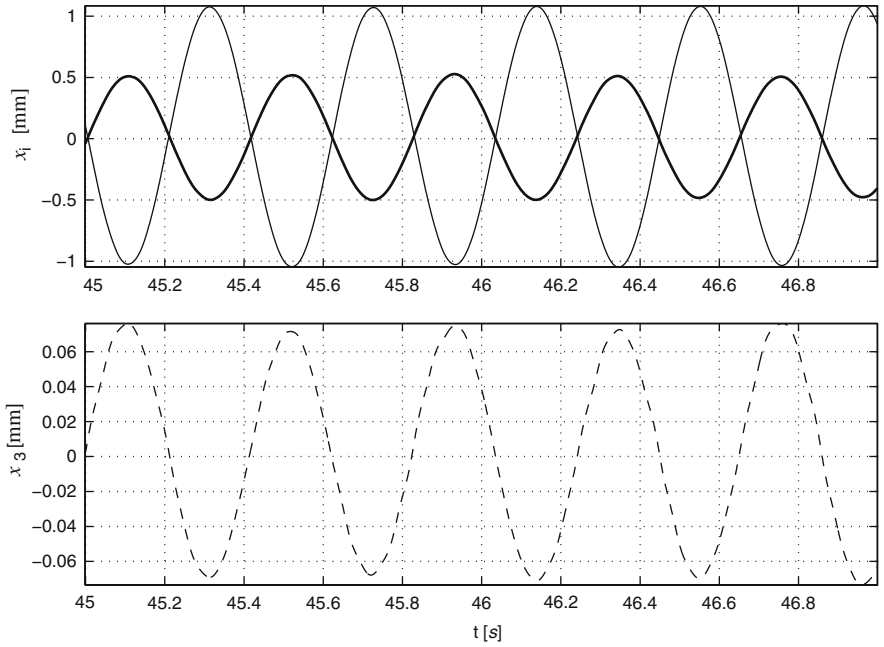


Fig. 8 Experimental results: Steady state behaviour of the system. *Top*: Displacement of the oscillators ($-x_1$, $-x_2$). *Bottom*: Displacement of the connecting beam

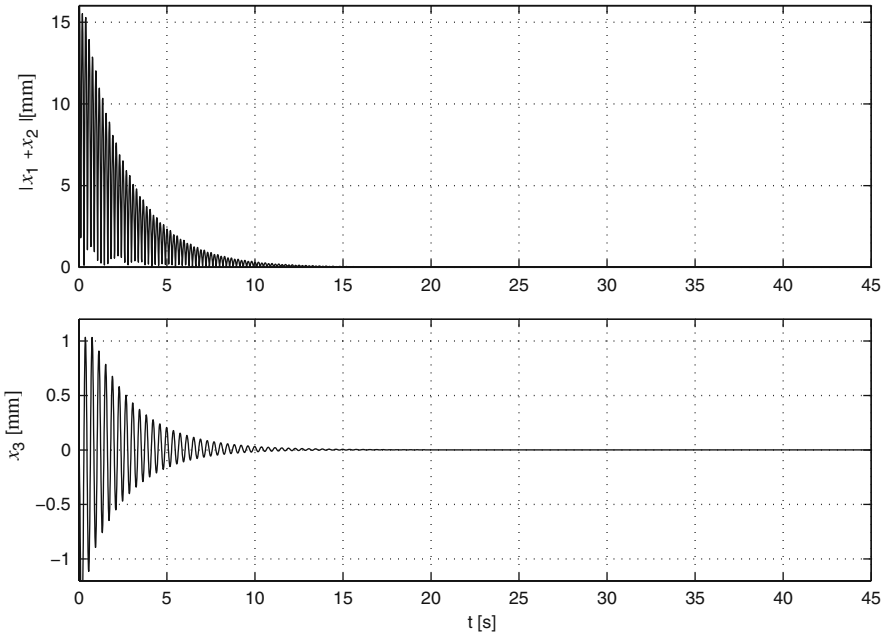


Fig. 9 Numerical results: (*Top*) Sum of the displacements of both oscillators. *Bottom*: Displacement of the connecting beam

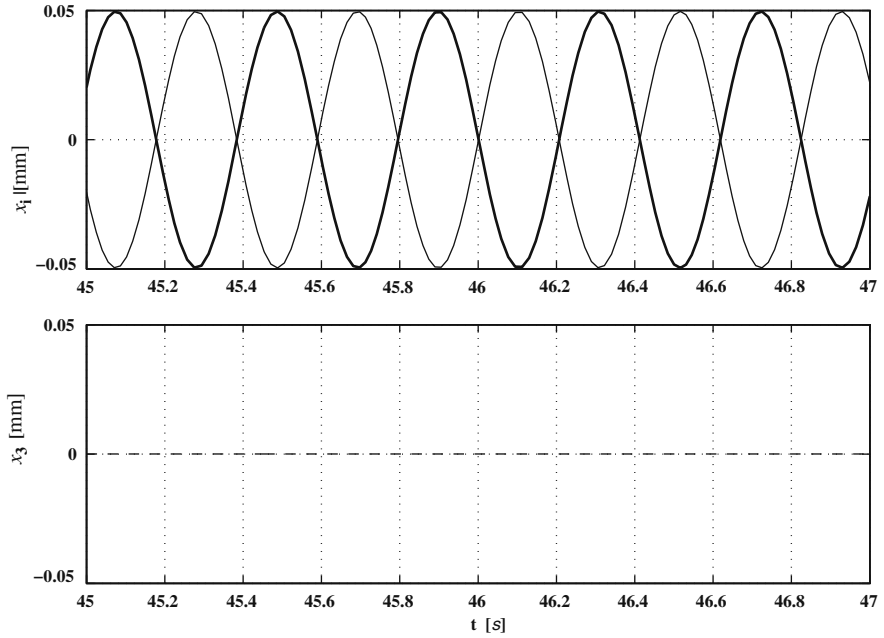


Fig. 10 Numerical results: Steady state behaviour of the system. *Top*: Displacement of the oscillators ($-x_1$, $-x_2$). *Bottom*: Displacement of the connecting beam

Table 2 Parameters in numerical simulation

$\omega_o = 15.26$	$\vartheta = 8.14$	$M = 0.8$
$m = 1$	$k = 1$	$b = 5$

6 Example 2: Two Coupled Rotary Disks

Next to the synchronization of Duffing oscillators we investigated synchronization in a system of coupled rotating discs as depicted in Fig. 11. First the dynamics of the system will be specified in more detail and next experimental results will be presented.

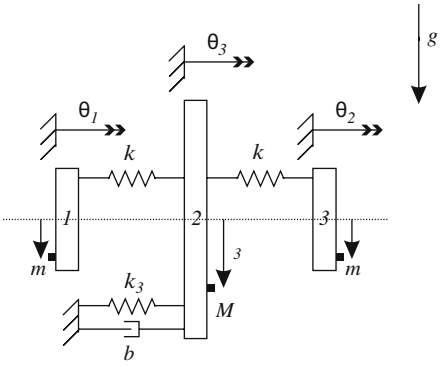


Fig. 11 Schematic representation of the set-up modeling two coupled rotating elements

6.1 Problem Statement

Consider the system as depicted in Fig. 11. This system consists of three discs. Discs 1, 2 represent the oscillators and disc 3 is connected to both other discs by torsion springs with stiffness k. Each of the discs has an eccentric mass at a distance ℓ_i from it's center ($\ell_1 = \ell_2 = \ell$). Furthermore the middle disc is coupled to the world by a torsion spring with stiffness k_3 and a torsion damper with constant b. The rotation of the discs is represented w.r.t. the world by the angles θ_i. The equations of motion of the system depicted in figure 11 are:

$$\ddot{\theta}_i = -\vartheta_i \left(k\left(\theta_i - \theta_3\right) + \delta_i \sin\theta_i\right), \; i = 1{,}2 \tag{19}$$

$$\ddot{\theta}_3 = \vartheta_3 \left(\sum_{j=1}^{2} k\left(\theta_j - \theta_3\right) - k_3\theta_3 - b_3\dot{\theta}_3 - \delta_3 \sin\theta_3 \right), \tag{20}$$

with $\vartheta_i = \frac{1}{m\ell_i^2 + J_i}$ and $\delta_i = m_i g \ell_i$. The modification to the set-up is now more involved than in the previous example. First of all, the translation coordinates x_i

should be mapped to rotation angles θ_i. Secondly, in case of the Duffing oscillator the actuation forces F_1 and F_2 are meant to act on both the oscillators and the connecting mass. In the situation depicted in Fig. 11 the actuation force generated to model the coupling between the oscillator discs and the middle disc by means of the torsion spring should again act on the oscillators and the connecting beam in our set-up. However, the part of the actuation force that models the influence of gravity on the oscillators should only act on the oscillators and not on the connecting beam, since in Fig. 11 the gravity on discs 1 and 2 exerts a force only on the corresponding disc and not directly on the middle mass.

In order to adjust the set-up in Fig. 5 to model the system in Fig. 11 the actuator forces are defined as:

$$F_i = \kappa_i(q_i) + \beta_i(\dot{q}_i) - \vartheta_i\,(\eta_i + g_i)\,, i = 1{,}2 \tag{21}$$

$$F_3 = \kappa_3(x_3) - \vartheta_3\,(\eta_3 + g_3) - \tilde{g}(\,\cdot\,), \tag{22}$$

with $\kappa_i(q_i)$ and $\beta_i(\dot{q}_i)$ as defined (4,5), $\eta_i = k\,(\theta_i - \theta_3)\,,\ i = 1{,}2$, $g_i = \delta_i \sin\theta_i$ and $\tilde{g} = \sum_{j=1}^{2} \vartheta_i g_i$. Damping is left to be the natural damping of the beam in the set-up. Furthermore, translation is mapped to rotation angles according to: $\theta_i = \frac{\pi}{2}\frac{x_i}{x_i^\star}$, with $x_i^\star$ is the maximal displacement of the oscillators and the beam, assuring $\pm\,90°$ turns in the rotation space.

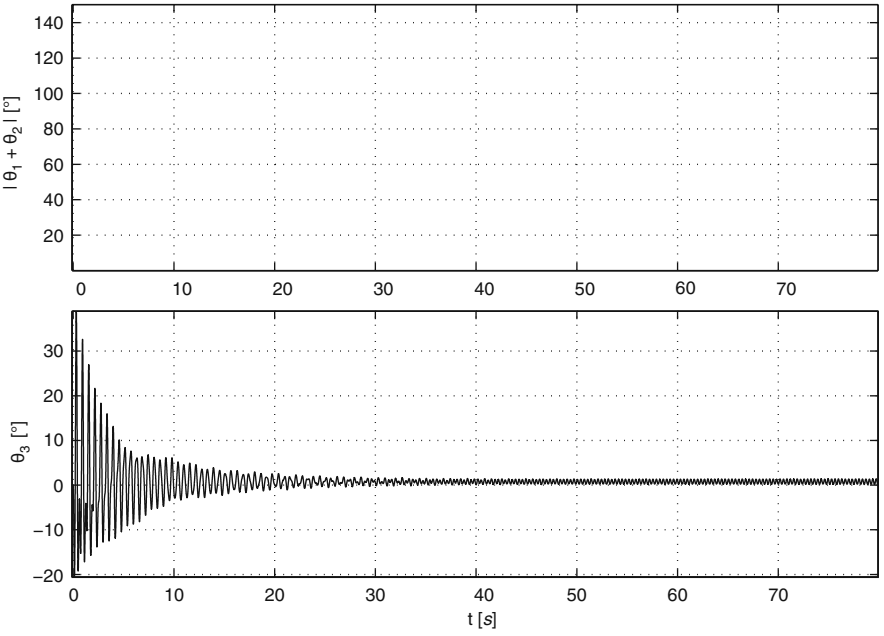


Fig. 12 Experimental results: (*Top*) Sum of the rotation angles of the outer discs. *Bottom*: Rotation angle of the connecting disc

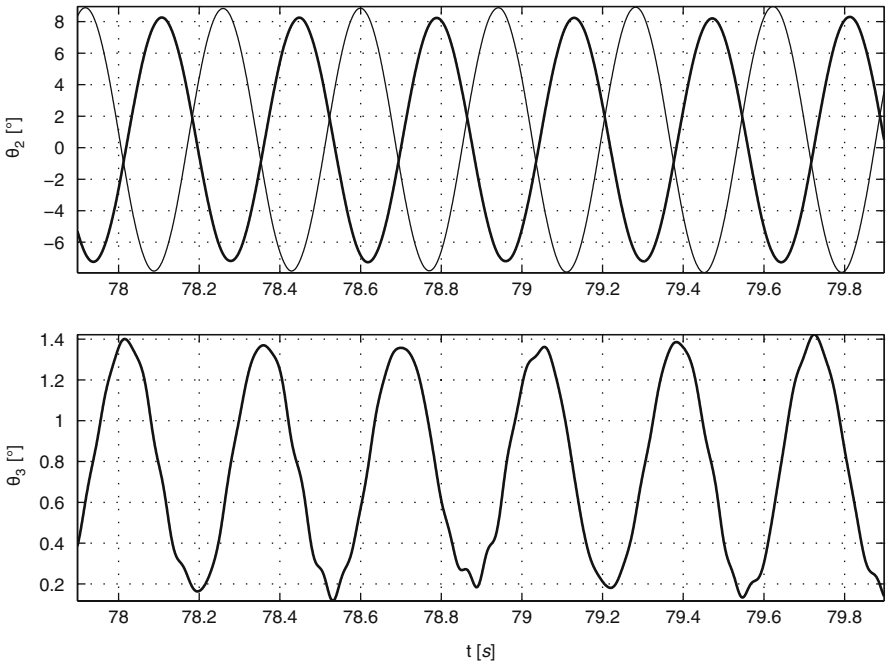


Fig. 13 Experimental results: Steady state behaviour of the system. *Top*: Outer discs ($-\theta_1$, $-\theta_2$). *Bottom*: Connecting disc

6.2 Experimental Results

Experimental results, are presented in Figs. 12 and 13. It becomes clear that approximate anti-phase synchronization occurs after about 20 s, like in the Huygens' pendulum set-up. Again complete synchronization does not occur because the oscillators are not identical. In addition Fig. 13 shows the steady state behaviour of the rotating system, from which the approximate anti-phase synchronized behaviour becomes immediately clear.

7 Conclusions

We presented a set-up capable of conducting synchronization experiments with a variety of different oscillators. Two sets of experimental results were provided that show the potential of this set-up. First we modeled and experimentally obtained synchronization between two coupled Duffing oscillators. Second, we showed that it is possible to model systems with rotating dynamics and to effectively model the local influence of gravity in this case.

In addition to studying uncontrolled synchronization the set-up has the potential to study controlled synchronization. Furthermore, we aim to model the Huygens

set-up and perform controlled and uncontrolled synchronization experiments with this type of dynamical system.

Acknowledgement This work was partially supported by the Dutch-Russian program on interdisciplinary mathematics "Dynamics and Control of Hybrid Mechanical Systems" (NWO grant 047.017.018).

References

1. C. Huygens. *Oeuvres compl'etes de Christiaan Huygens*. Martinus Nijhoff, The Hague, vol. 5, 17, pp. 241–262, 156–189 (1893, 1932).
2. M. Bennett, M. Schat, H. Rockwood, and K.Wiesenfeld. Huygens's clocks. *Proc. R. Soc. A Math. Phys. Eng. Sci*., **458**(2019), 563–579 (2002).
3. J. Pantaleone. Synchronization of metronomes. *Am. J. Phys*., **70**(10), 992–1000 (2002).
4. M. Senator. Synchronization of two coupled escapement-driven pendulum clocks. *J. Sound Vib*., **291**(3–5), 566–603 (2006).
5. N. V. Kuznetsov, G. A. Leonov, H. Nijmeijer, and A. Yu. Pogromsky. Synchronization of two metronomes. *3rd IFAC Workshop "Periodic Control Systems"*. St. Petersburg, Russia, CDROM (2007).
6. W. Oud, H. Nijmeijer, and A. Yu. Pogromsky. A study of Huygens synchronization. Experimental results. *Proceedings of the 1st IFAC Conference on Analysis and Control of Chaotic Systems*. France, CDROM (2006).
7. A. Yu. Pogromsky, V. N. Belykh, and H. Nijmeijer. Controlled synchronization of pendula. *Proceedings 42rd IEEE Conference on Decision and Control*. Maui, Hawaii, pp. 4381–4368 (2003).
8. A. Yu. Pogromsky, V. N. Belykh, and H. Nijmeijer. A study of controlled synchronization of Huygens' pendula. In K. Y. Pettersen, J. T. Gravdahl, and H. Nijmeijer, editors, *Group Coordination and Cooperative Control*, Lecture Notes in Control and Information Sciences. Springer, New York, pp. 205–216 (2006).
9. S. Strogatz. *SYNC. The Engineering Science of Spontaneous Order*. Hyperion, New York (2003).
10. A. Pikosvky, M. Rosenblum, and J. Kurths. *Synchronization*. Cambridge University Press, Cambridge (2001).
11. J. Buck. Synchronous rhythmic flashing of fireflies. II. *Q. Rev. Biol*., **63**(3), 265–289 (1988).
12. I. I. Blekhman. *Synchronization in Science and Technology*. ASME Press, New York (1988).
13. W. Lindsey. *Synchronization Systems in Communications and Control*. Prentice-Hall, Englewood Cliffs, NJ (1972).
14. M. H. L. M. van den Tillaart. *Design of a mechanical synchronizing system for research and demonstration purooses for D&C*. Master Thesis, Eindhoven University of Technology (2006).

Controlling Chaos: The OGY Method, Its Use in Mechanics, and an Alternative Unified Framework for Control of Non-regular Dynamics

G. Rega, S. Lenci, and J.M.T. Thompson

Abstract In this chapter we review the development of the control of chaos theory subsequent to the seminal paper by Ott, Grebogi and Yorke in 1990. After summarizing the main characteristics of the OGY method, we analyze and discuss various applications in several fields of mechanics. We then illustrate the main aspects of an alternative control method which aims at controlling the overall system dynamics instead of stabilizing a single periodic orbit, as the OGY method does. The two methods are both based on the modern idea of exploiting the chaotic properties of systems, instead of simply eliminating chaos.

This paper is one of a collection written in honour of Celso Grebogi, on the occasion of his 60th birthday. So we have thought it appropriate to start with short personal reminiscences by two of the present authors.

Some personal remarks by Michael Thompson.
Thinking back to my first encounters with chaos in the early 1980's, I remember that the work of "Grebogi, Ott and Yorke" soon became indelibly etched on my mind. Interested by some unexpected subharmonic resonances that had been observed when oil tankers were moored to articulated off-shore towers, I had modelled the system as a driven impact oscillator. This exhibited a lot of fascinating nonlinear phenomena, including what David Rand (at Warwick University) told me were chaotic motions: about which I had to do some very rapid reading! My first publication about the chaos in this impacting system [1] appeared in 1982 which was, interestingly, the year of Celso's first major paper on chaos [2]. The latter, authored by Grebogi, Ott and Yorke, was a seminal paper dealing with the bifurcational crises of chaotic attractors: to date it has attracted 483 citations. It became one of my standard references in, for example, my paper on the design of compliant off-shore structures [3].

Having linked up with Bruce Stewart at the Brookhaven National Laboratory, New York, our book [4] entitled *Nonlinear Dynamics and Chaos* (Thompson and

G. Rega (✉)
Dipartimento di Ingegneria Strutturale e Geotecnica, Sapienza Universitá di Roma, I-00197 Rome, Italy
e-mail: giuseppe.rega@uniromal.it

M. Thiel et al. (eds.), *Nonlinear Dynamics and Chaos: Advances and Perspectives,* Understanding Complex Systems, DOI 10.1007/978-3-642-04629-2_11,

Stewart, published by John Wiley) appeared in 1986. Key references at the end of this book included Celso's second paper on crises [5] (838 cites)), and his work on quasi-periodicity of order three [6] (76 cites).

I next turned my attention to the chaotic phenomena that triggered the escape from a potential well, publishing in 1989 a paper in the *Proceedings of the Royal Society* [7] which itself has attracted 120 citations. In this work I found it necessary to reference six papers by Celso including [5] described above. A paper of particular significance to me, by McDonald et al. [8] (351 cites), was on fractal basin boundaries which I warned could seriously decrease the integrity of engineering systems.

From this spectacular start, Celso's career has gone from strength to strength, and he is now on the distinguished ISI List of Highly Cited Researchers. Meanwhile the writers of the present paper note with special pleasure and satisfaction that the paper entitled *Controlling Chaos* [9] published by Ott, Grebogi and Yorke in 1990 has been selected by the American Physical Society as one of the milestone pieces of research from the last 50 years. This puts Celso (alongside many Nobel Prize winners) in the list of milestone contributors which has been created by the society to celebrate the 50 years of *Physical Review Letters*. Not surprisingly, this paper of 1990 is Celso's most highly referenced work, having been cited 2,711 times.

After seeing his early papers in print, I soon enjoyed meeting Celso at many international conferences throughout the world: and now, I am delighted and honoured to be getting to know Celso and his family very well indeed because we both hold Sixth Century Chairs at the University of Aberdeen (his full-time, mine part-time).

Some personal remarks by Giuseppe Rega.
I first met Celso Grebogi at the IUTAM Symposium on *Nonlinearity and Chaos in Engineering Dynamics* organized by Michael Thompson at University College London, 1993. Within a top level specialty meeting of scientists from engineering mechanics working on nonlinear dynamics and their applications, Celso was invited, from a different environment, to mark a conscious and widespread transition from the traditional asymptotic/numeric treatment of nonlinear dynamics problems followed within the community of mechanicians, to a more modern and integrated approach relying additionally on geometric and experimental methods. Celso fully succeeded in his own task, by providing the attendees with a clear feeling of the importance of a dynamical systems approach to nonlinear/chaotic dynamics in engineering and by introducing the newly developed concept and technique of chaos control. Since then Celso has become for the engineering community the most appropriate and acclaimed scientist for knowledge transfer from Physics as well as for cross-fertilization between Physics and Engineering.

Within his extraordinary number of achievements on cutting edge issues in dynamical systems theory and applications and assuming an engineering mechanics scholar's viewpoint, I just mention two items. (i) All dynamical systems scientists from engineering have become familiar with concepts like sudden change bifurcations in chaotic attractors and basin boundary metamorphoses by reading Celso and co-authors' papers; (ii) the OGY method paved the way for a modern and

comprehensive notion of control of chaos, including its suppression, enhancement or use, which is of major importance for practical applications.

In the last decade, Celso's interests meaningfully spread to include problems of dissipative dynamics typically occurring in engineering applications. His former interests towards applications of chaos and control to, e.g., plasma and laser systems or communication, evolved to include such typical systems and problems in mechanics as pendula, ship capsizing, impact and dry friction oscillators, fluidized beds, hydrodynamical flows, suspension bridge models, and spacecraft steering.

Within the scientifically fertile and prestigious Maryland group, Celso gave fundamental contribution as regards both the scientific aspects and the worldwide promotion of knowledge on chaotic dynamics and their applications. His broadmindedness and capability in cross-disciplinary relationships allowed him to play a role of most visible and internationally recognized scientist by neighbouring scientific communities dealing with complexity. A capability to lead and motivate research groups and disseminate knowledge to companion communities and wider non-specialists' audiences that were fully confirmed when he moved to the University of Sao Paulo and, more recently, to the University of Aberdeen.

1 Controlling Chaos: A Hot Topic at the Change of the Millennium

Among various other research topics which attracted the attention of researchers in the 20 years across the change of the Millennium, control of nonlinear dynamics and chaos is certainly one of the most important and well identified, developed at lenght from a few seminal ideas and methods, and yet far from being exhausted.

The quality of papers published on this topic cannot be judged by numbers or quantitative indicators, and indeed we found, as in other research fields, both very good – seminal, indeed – and poor works. However, numbers provide an understanding on the "impact" of this discipline on the advancement of science, engineering, practice and, more generally, of knowledge. Thus, in the certainly incomplete list of references we quote:

- 8 books [10–17] by the most important publishers;
- 9 journal special issues [18–26] of the most renowned scientific journals;
- 13 review or survey papers [27–39] in leading international scientific journals.

Even more impressive is the number of articles and conference papers published on the subject. According to "Scopus" (`www.scopus.com`), looking for "control of chaos" in the title, abstract or keywords, on 3 February 2009 we got 7,526 papers, with a growing rate of papers published per year. This extends the data of Fradkov et al. [33] who refer to 2,700 papers published in peer-reviewed journals up to the year 2000, with more than a half in 1997–2000.

One interesting aspect of the "control of chaos" discipline is the cross-disciplinarity, which is a consequence of the cross-disciplinarity of its parent, the "chaos theory." Thus, we have papers from mathematics, physics, chemistry,

biology, medicine, economics, and various fields of engineering (see Sect. 3 for more details related to mechanical engineering and structural dynamics).

The starting point of the theory and applications of "chaos control" is unanimously considered to be the work of Ott, Grebogi and Yorke [9]. Actually, other works aimed at "suppressing chaos" in some systems appeared at about the same time or earlier (e.g., [40, 41]), but the seminal idea of exploiting the chaotic behaviour of systems in order to control their dynamics was first undoubtedly presented in the OGY paper. It represents a major improvement with respect to other points of view on the matter, independently pursued within different scientific communities (of mathematicians, physicists and engineers): in fact, it marks the fundamental passage from analysis to synthesis of chaotic properties, just based on the knowledge and exploitation of the dynamical systems theory deeply studied in the past.

Since then, control of nonlinear dynamics and chaos – intended in a broad sense – has become a hot trans-disciplinary research topic, as shown not only by the already mentioned great number of scientific publications but also by its increasing popularity within the environment of non-technical chaos connoisseurs.

In most of the works on the subject, the 1990 OGY paper is quoted along with some of the articles, aimed at improving and discussing the original idea or pursuing possible relevant applications, successively published by the same group of authors, with further collaborators, or by independent scientists. Without any attempt at completeness, one can mention, e.g., [42–47], in the first group, and [48–53] in the second group.

Yet, as already said, various underlying notions of chaos control stay in the background of the research activity in the field. In line with some first attempts, in their papers several authors aim at merely *removing* chaos, by means of classical control techniques [54], empirical methods [55], or other clever approaches [56]. These works are also referred to as "control of chaos", although in this matter some authors, and we agree with them, prefer the name "suppression of chaos" (e.g., [36, 57–59]), which focuses on the effects of control rather than on the underlying skill of the control method. Thus, today, one refers to chaos control both when

1. chaotic transients and/or attractors are eliminated "tout court," even if the tools employed to eliminate chaos have nothing to do with it, and when
2. typical properties of chaotic dynamics are involved in the control process, irrespective of the actual tools being employed to control the system and, indeed, of its actual response.

This twofold control aspect is already clearly highlighted in Linder and Ditto [36], when they assert that "... some techniques merely suppress or remove chaotic behavior ... others actually exploit chaotic behavior ...".

Within the latter, more general and modern, perspective, the capability to exploit typical properties of dynamical systems undergoing chaotic behaviours in order to *control* the system response actually refers to either *suppressing* or *enhancing* chaos (the so called anti-control of chaos), or even *using* it, based on the

specific goal of the considered technique, on the system at hand, and on the area of interest [38].

Another basic issue is concerned with whether the dynamical phenomenon to be controlled by whatever "chaos control" technique has to be intended as "chaos" in a strict sense or, rather, as any kind of *complex behaviour* of a dynamical system, which may have different aspects according to whether a theoretical or a practical viewpoint is adopted. In the former, dynamical systems oriented, perspective, one can think of, e.g., any global bifurcational event possibly entailing complex behaviour of the system, of the relevant escape from a safe subset (a potential well or a basin of attraction) in parameter control space, or of the synchronization of different oscillators. In the latter, application-oriented, perspective, the kind and meaning of the "complex" phenomenon of interest is dictated by the practical goal to be attained.

In view of this extended notion of chaos control, no relevant exhaustive *classification* seems to be yet available, although some earlier interesting attempts have been made. For example, Chen and Dong [10] proposed a classification based on various tools employed in the control process (parameter-dependent approaches, open-loop strategies, engineering feedback control, adaptive control, intelligent control, etc.), while Fradkov's "Chaos Control Bibliography (1997–2000)" [33] also contains a classification of the various applications in science and engineering. Recent classifications basically distinguish between feedback (or close-loop) and non-feedback (or open-loop) control techniques.

A more phenomenologically based classification relies on the ascertainment of how one can statically or dynamically modify either the system parameters – which represents the most basic form of passive control at the design stage – or the excitation, to attain the control goal. In particular,

1. the OGY's method [9],
2. some classical methods (CM) of control theory [60],
3. the "control by system design" (CSD) [28],
4. the "parametric variation methods" (PVM) [10],

aim at somehow modifying the system. Note that also the OGY method could be classified as a PVM, but it has some specific features and we prefer to keep it separate.

In turn, methods acting on the excitation include

1. classical methods where a properly modified input (periodic or aperiodic, open-loop or feedback) is applied to the system (CM) [54, 56, 60, 61];
2. the "control through operating conditions" (COC) [28] based on modifying the frequency and/or the amplitude of the excitation;
3. the methods based on either combining parametric and external excitations (PEE) [41], or applying weak periodic perturbations (WPP) [17], or modifying the shape of the excitation (SE) [40, 62, 63].

Obviously, different methods are expected to give different performances, at least theoretically. Thus, one has methods aimed at:

1. stabilizing an unstable zone of parameter space (CSD, COC, PVM);
2. moving away from (previously known) chaotic zones (CSD, COC);
3. stabilizing a given, erratic solution (CM, OGY);
4. overall regularizing the system dynamics, irrespective of single solution behaviour (PEE, WPP, SE).

It is the authors' opinion that these latter distinctions are very important from a practical point of view because they suggest the use of the most appropriate control method fitting prescribed technical requirements.

Focusing the attention on the exploitation of chaotic behavior, there is even another possible classification, which seems not pursued in the literature, where various methods can be grouped according to the chaotic properties involved in the control. In this respect, one can distinguish methods based on

1. the properties of the saddles embedded in the chaotic attractor (OGY);
2. the ergodicity of the chaotic attractor (for example, all methods where a preliminary targeting step [46, 64, 65] is followed by the application of control tools in the neighborhood of the chosen control area belong to this class);
3. the sensitivity to initial conditions [45];
4. the occurrence of homo/heteroclinic bifurcations (PEE, WPP, SE).

Besides the formulation of different techniques for chaos control, the nearly two decades straddling the change of the Millennium have seen also a large number of relevant *applications* to mathematical, physical, biological, and engineering systems, within different technical fields. Of course, no attempt is made herein to comprehensively report on them. In contrast, consistent with one of the goals of this chapter, attention is focused on applications of chaos control in *mechanics* although, at a first sight, they could appear relatively few and of only minor importance.

Up to the authors' knowledge, earlier review papers explicitly devoted to this topic date back to 1993 [28]. The authors report on the newly developed (at that time) OGY method, and propose to control a mechanical system by either changing the forcing characteristics (e.g. the excitation frequency) – with the aim of moving away from a previously known chaotic region – or modifying any system property, such as inertia or stiffness. All of the proposed techniques fall in the area of "suppressing" chaos, and while being illustrated with reference to a Duffing oscillator (possibly equipped with also a tuned mass damper), they do not actually refer to specific issues associated with the mechanical nature of the systems.

A recent survey of the application of various methods of chaos control to mechanical systems [35] reports quite a long list of mechanical systems or processes for which control of nonlinear dynamics and chaos has been addressed in the literature, furnishing at least one reference for each of them. The list includes pendulums, beams and plates, systems with friction and/or impacts, spacecraft, vibroformers, microcantilevers, ship oscillations, tachometers, rate gyros, Duffing oscillators, robot-manipulator arms, earthquake civil engineering, milling processes, whirling motions under mechanical resonance, and systems with clearance. The items in the list highlight the quite scattered nature of the work in terms of the *mechanical*

complexity and variety of the systems, of the involved dynamical processes, and of specific control goals.

Indeed, choosing a proper framework within which to organize, present and discuss the state of art in the field is not an easy task because, depending on various application fields, mechanics is concerned (i) with both discrete (finite-dimensional) and continuous (infinite-dimensional) dynamical systems, (ii) with their possibly reliable low-dimensional models, and (iii) with a considerable richness of dynamical processes and phenomena to be possibly controlled. Though being mechanics mostly concerned with large scale systems traditionally analyzed and designed via linear techniques, a large amount of research made in the last 30 years has highlighted a cornucopia of nonlinear dynamic phenomena, including chaotic responses, along with their importance in mechanical and structural applications. Yet, regular nonlinear phenomena of such a variety of systems are already quite complicated in themselves, and as such exhibit a major importance in the real world of mechanics. As a matter of fact, the question of how important and pervasive chaotic phenomena are in the behavior of mechanical/structural systems, is still to be answered. Accordingly, it is quite difficult to select studies devoted to chaos control in mechanics in a strict sense, whereas the topic has to be intended in the wider sense of control of *dynamical complexit*y, and is often addressed in the literature within the more general area of control of nonlinear, wanted or unwanted, phenomena.

On the other hand, the inherent complexity of systems from mechanical science and engineering poses the challenging problem of identifying proper reduced order models able to reliably describe the ensuing rich nonlinear dynamics; an issue which becomes even more demanding if one is interested in dealing with possibly complex nonregular response of the actual system and with its control. This basically entails one major consequence and one question.

1. A large majority of research works on control of chaos in mechanics is concerned with (relatively simple) archetypal nonlinear oscillators often representing idealizations of more involved discrete systems encountered in real mechanical engineering or minimal discretized representations – according to some reduction technique – of the infinite-dimensional continuous systems typical of structural engineering. This also entails that in most research papers the interest is focused on the control technique – or on other dynamical or control aspects transversal and unifying with respect to other scientific/technological fields – rather than on the actual mechanical peculiarities of the considered systems.
2. Even being convinced of the significance of nonregular dynamics – and of their control – for real mechanical/structural systems, how representative of their actual behavior are the chaotic regimes highlighted for such reference archetypal oscillators? Of course, this is a general point arising in any modeling and reduction problem, but it is felt to have a special meaning in mechanics owing to the above mentioned inherent complexity of the involved large-scale systems. The only possibility to clarify this issue consists in highlighting the actual occurrence, and the features, of chaotic response in experimental (mechanical and structural) systems, in developing refined theoretical and numerical models of the actual

system, and in cross-validating the two approaches. In this respect, considerable research is going on to highlight occurrence and features of chaos in experimental and/or refined theoretical models, but relatively little is yet done as regards their control.

Altogether, considering such a rich and intriguing framework which encompasses several aspects, the objective and organization of this chapter can be stated as follows.

The major role played by the OGY method in establishing the basic philosophy of chaos control is recognized in Sect. 2, along with its improvements, extensions and applications.

The use of OGY method for chaos control in mechanics is systematically addressed in Sect. 3, by identifying distinguishing classes of systems and applications, and by discussing them in the framework both of the capabilities and of some limitations of the method.

Then, a substantially alternative technique – according to the previously mentioned classification criteria – is presented in Sect. 4. It consists of an overall procedure for controlling the non-regular dynamics of systems triggered by the occurrence of homoclinic or heteroclinic bifurcations, which is discussed in its generic/non-generic features and with regard to a number of applications to systems of interest in nonlinear structural dynamics and engineering.

2 The Paradigmatic OGY Method for Chaos Control

The OGY method was first formulated in the celebrated work by Ott et al. [9]. A web-based search on "Scopus" looking back to the year 1998 (i.e., only 10 years back from now, instead of 18) already furnishes more than 2,300 articles citing the original OGY paper, with many of them also citing various other papers concerned with the method and published in the following years. In fact, the scientific literature in the area registered the publication of a great number of papers, co-authored either by one of the original authors with collaborators or by other scientists somehow related with the University of Maryland Chaos Group, for the whole decade of the Nineties. These papers provided explanations, clarifications, extensions and improvements on the theoretical and computational aspects of the method, as well as a great number of experimental implementations, applications in a wide variety of technical fields – ranging from mechanical systems to electronics, laser technology, chemical reactions, heart tissues, communications –, and modifications/refinements aimed at overcoming some observed limitations or drawbacks. The number of papers ensuing from the original one is so large that it makes no sense to list or report on them. A similar reasoning can be done as regards the theoretical and practical aspects of the method, which are somehow summarized in most – if not all – of the articles subsequently published by different authors and concerned with OGY control of chaos.

Quite soon the method became popular also within the less strictly technical literature, as witnessed by a number of articles appeared, e.g., in Nature [47], Nonlinear Science Today [66] and Physics Today [67]. In both the hard and soft science respect, one could even run the risk to somehow compare, though to a lower scale, the paradigmatic role played by the OGY method in the area of control of chaos at the round of the Millennium with the revolutionary role played by chaos in the area of nonlinear science in the second half of the XX century!

Based on previous points, no analytical description of how the method works is reported herein. Yet, to the sole purpose of properly frameworking the discussion on the use of OGY control of chaos in mechanics, successively made in Sect. 3, a brief summary of the main qualitative features of the method is provided following the updated – though popular – exposition given in Ott [68].

Two fundamental aspects of chaos are:

1. The exponential sensitivity of orbits to small perturbations.
2. A very complex orbit structure making possible many different motions in the same system at fixed parameter values.

Actually, these are not independent properties, but rather two sides of the same "chaos coin". The first property, typically quantified by the largest Lyapunov exponent, refers to the difficulty in the prediction of sufficiently distant future states of the system, while the second property, most often quantified by an entropy measure, refers to the fact that chaotic attractors often have embedded within them an infinite, dense set of unstable periodic orbits. While chaos is commonly viewed as undesirable, these properties can also be of great benefit in certain situations where one wishes to control the system dynamics. In particular, if the system is sufficiently well characterized and noise is not too large, then the first property implies that relatively large changes in the eventual location of orbit points can be made using only small changes in a control variable, while the second property implies that the dynamical changes so produced may be very diverse, leading to a situation with a great deal of flexibility. The two methods that are useful in utilizing chaos in an effective way are

1. *control*, namely the feedback stabilization of selected unstable orbits embedded in the attractor, and
2. *targeting*, namely the rapid steering of a chaotic orbit to the vicinity of a desired point in phase space.

Here *control* means *feedback control*, i.e., measurements of the state of the system are regularly taken, and, based on them, some controllable parameter (or set of parameters) is adjusted so as to achieve some goal. The desirable goals may vary, and different types of goals lead to qualitatively different control problems. Apart from those of interest for specific applications, general goals include:

1. Given a steadily running chaotic system, how can one improve its time averaged performance? Here performance is defined with reference to the specific function the system is meant to carry out.

2. Given a chaotic system in a given state at some specific time, how can one nudge the subsequent orbit to evolve rapidly from its current state to a different target location in state space ?

For the first goal, the complex orbit structure is most relevant; for the second goal, exponential sensitivity is most relevant. In all cases, the fundamental attributes of chaos imply that the control goals can be potentially achieved by use of only small controlling perturbations. Thus, one consequence of chaos is that control can be accomplished with low-energy/low-force controllers.

In order to see how this might be done, one first refers to the infinite number of unstable periodic orbits (UPOs) embedded within typical chaotic attractors. If one had the ability to perfectly place an initial condition on any chosen one of these UPOs, then an infinite number of different types of motion could be achieved. Moreover, if no noise were present in the system, such motions would ideally continue indefinitely in time.

The performance of a system, in general, depends on its state and its history. Let $x(t)$ denote the system state as a function of time t. Assume that the performance P of some steadily running process can be given as the time average of some quantity f that is a function of the system state $P = \langle f(x(t)) \rangle$, where the angle brackets denote a time average. For each UPO, a different state trajectory $x_i(t)$ results, where the subscript i labels the particular UPO. Consequently, each unstable periodic orbit i will have associated with it a performance $P_i = \langle f(x_i(t)) \rangle$, and these performance values will typically be different for each i. Moreover, it can be shown that, under suitable conditions, the performance P_c for the chaotic orbit is a weighted average of the performances P_i attained by the periodic orbits. The implication of this is that some of the P_i will be larger than P_c. If the system can be controlled to such a periodic orbit, then the system performance will be improved.

The next question is how to control the system so that it follows the chosen UPO rather than the chaotic orbit $x_c(t)$. One way of doing this is to wait until the ergodic uncontrolled chaotic orbit $x_c(t)$ comes close to the desired UPO, and then give it a small kick to place it on or very near the UPO. Due to several factors (system noise, our inability to kick the orbit to a precise location, an imperfect knowledge of precisely where the desired orbit lies), the kicked orbit will not be exactly on the desired UPO. Since the UPO is unstable, the system orbit will begin to move away from it. As soon as this is discerned, a small kick can be reapplied to reposition the system orbit closer to the desired UPO. By doing this continually, the orbit can be kept close to the desired UPO indefinitely. Note that for small noise and inaccuracy, the size of the kicks required to maintain the system orbit near the desired UPO is correspondingly small, approaching zero for the noiseless, absolutely accurate case. Thus, in many situations, the goal can be achieved with small controls. In comparison, one should note that in other proportional feedback control schemes where the targeted state might not be a part of the intrinsic dynamics, large controlling signals might be needed even if the system is noiseless.

Two issues that must be considered for such implementations are

1. how to determine and locate UPOs embedded in a chaotic attractor, and
2. how to make the small controlling kicks.

Determining UPOs If a very accurate analytical model of the system is available, then standard numerical techniques (e.g., Newton's method applied to the fixed point equation of the n times iterated map) can be applied to determine UPOs. However, in many cases of experimental interest, an analytical model of the system under study may not be available. In such cases, it is still possible to determine UPOs purely from data by recording the trajectory of a free-running (i.e., uncontrolled) chaotic orbit. The idea is to use state space embedding and attractors reconstruction techniques [69]. Techniques for finding UPOs from data have been discussed by So et al. [70] and by Pierson and Moss [71], among others.

Another issue is how many UPOs on the chaotic attractor need to be determined. Hunt and Ott [72] showed that it is seldom necessary to determine a very large number of UPOs since maximal performances typically occur on low period UPOs, which are limited in number.

UPO Controlling Algorithms Having chosen a suitable UPO embedded within the attractor, it remains to specify how the small controls should be programmed to maintain the system orbit on the chosen UPO. Several ideas have been proposed to accomplish this:

- The original OGY technique employed in Ott et al. [9] is to use the control to place the orbit on or near the stable manifold of the desired UPO.
- Romeiras et al. [73] discuss use of the "pole-placement" technique, which is standard in control theory.
- Dressler and Nitsche [74] and So and Ott [75] show how UPO control can be implemented using only time-series measurements of a single scalar state variable.
- Socolar et al. [76] present a technique particularly useful for the control of very fast dynamics.

Targeting refers to the control goal of quickly bringing an orbit to some desired location in state space. The basic idea is that since chaos is exponentially sensitive to small orbit perturbations, such orbit perturbations become large in a relatively short time. Moreover, if the perturbations are very carefully chosen, then there is the hope that the orbit can be efficiently directed to the target using only small controls. Several techniques for achieving this goal have been formulated [46, 77–79].

A number of improvements and modifications of the OGY control, aimed at overcoming some of the associated limitations, will be mentioned in the following section in connection with use of the method, or of any of its many variants, for control of chaos in mechanical systems. It will be shown how the method works well, both numerically and experimentally, for a considerable number of relatively simple systems and models. Yet, one major problem is concerned with the extent of the impact that the method can have for chaos control in real world applications.

This is linked with the general problem, mentioned in Sect. 1, of how much reliably representative of actual systems behaviour are the nonlinear dynamic phenomena and their control as evidenced for substantially archetypal models: in this sense, all possible limitations, if any, are not to be abscribed to the sole OGY method.

3 Use of OGY Method for Control of Chaos in Mechanics

Two features of a chaotic system make it a candidate for OGY control [80]: (i) the presence within the strange attractor of one or more unstable fixed points of the saddle-node type, to be possibly stabilized, and (ii) an accessible system parameter through which an applied control input changes the location of the unstable fixed point in the phase plane, adjusting it in such a way to ensure that the trajectory returns to the vicinity of the original one.

In theory, the OGY method is a "black box" technique that can be applied to any system possessing these two attributes, even to experimental systems for which no accurate analytical or numerical model exists. Indeed, it is not necessary to have a mathematical model to achieve the control goal since all control parameters may be resolved from time series analysis with the assistance of delay coordinate embedding techniques [81, 82]. Yet, the method does not seem to work well when a relatively large influence of noise exists as in usual experimental applications (see [83] for a feedback controlled pendulum), although it has been applied successfully to several experimental models [47].

Also, some features of the method, such as considering a local linear return map in the region around the unstable fixed point to be stabilized and applying small control forces to ensure that the oscillator continually returns to the region of phase space described by such a map, may render the control ineffective or impractical for some sensitive systems. This can occur, e.g., if the true dynamics change too quickly in the region of interest, if the region of linear validity is too small for some unstable fixed points, or if the unstable eigenvalue is so large that points near the fixed point are swept away very quickly and tend not to return to the linear region [84]. Moreover, since control only begins when the trajectory enters for the first time the region where the linear map applies, it can happen that for some chaotic systems this may take thousands of cycles, depending on the effect of the control parameter on the estimated location of the perturbed fixed point and on the size of the linear capture region. So, the OGY method, which is locally linear, may not be practical for all chaotic systems.

Another issue related to control of chaotic attractors is the existence of a sometimes long chaotic transient before the system settles onto the stabilized orbit. This transient arises because time-dependent parameters can only be effectively applied to stabilize an orbit when the system falls within a narrow region around it. For typical initial conditions the system wanders chaotically and since the attractor is ergodic it will eventually fall sufficiently close to the fixed point to be stabilized. The length of the chaotic transient for randomly chosen initial values has

an exponential distribution. To reduce the necessary time for stabilizing the target orbit, the OGY method was modified by Shinbrot et al. [46] (SOGY method) by exploiting the extreme sensitivity of a chaotic attractor to tiny perturbations in initial conditions to rapidly direct a system from a given initial state to a desired one. Overall, this targeting approach provides a way of directing trajectories from any initial conditions to the desired fixed point, and is shown to be quite successful.

There are other improvements of the OGY method aimed at overcoming some of its original limitations, as for example control of high periodic and high unstable UPO [85, 86] and control using time delay coordinates [74, 75, 87]. In this respect, it is worth mentioning how many of the presently available techniques aiming at controlling chaos by stabilizing UPOs represent some interpretation or extension of the original OGY method.

3.1 The Pendulum System

The mathematical pendulum is the archetypal nonlinear oscillator in mechanics. Accordingly, it has been considered for validating chaos control techniques, testing their effectiveness and comparing them. As regards OGY-based control methods, besides several numerical applications, also a number of experimental applications have been concerned with nonlinear pendulums [85, 87–89], as well as a double pendulum [90].

Starrett and Tagg [88] controlled chaos in a pendulum with vertically driven pivot over a range of driving frequencies and damping levels by using two variations on the OGY method, namely by proportional adjustments of damping for fixed time intervals or by proportional time intervals for fixed levels of damping, and with the eddy-current damping as control parameter.

Baker [91] studied the control of a chaotic damped driven pendulum (with one external excitation) by using OGY method for stabilizing an UPO through a feedback mechanism that periodically adjusts the damping parameter of the pendulum.

A number of papers/applications used some kind of "OGY-like" methods. For example, based on an optimal control algorithm "belonging to the parameter variation techniques introduced in [9]" and minimizing the distance between a chaotic trajectory and a desired periodic orbit [92], Bishop et al. [93] realized flexible control of tumbling chaos in a parametrically excited pendulum by stabilizing it onto a variety of oscillating or rotating periodic states, located via coupled symbolic-numerical dynamics, by small adjustments of the driving frequency.

Several authors compared the capability and efficiency of various techniques by referring to the (mathematical) pendulum, with the interest being focused rather on the former than on the latter.

Yagasaki and co-workers investigated chaos control for a pendulum subjected to feedforward and feedback control. Yagasaki and Uozumi [89] showed that the chaotic dynamics resulting from transverse intersection between the stable and unstable manifolds can be stabilized to the target saddle-type periodic orbit by

using OGY and SOGY methods. Since the former method required a very long time for stabilization of the targets and the latter was not effective when modeling errors existed, Yagasaki and Uozumi [48] developed another approach where nonlinear approximations are used for the chaotic dynamical system and for the stable manifold of the target. It also includes the OGY method as a special case and is no more than just the pole placement technique. Yagasaki and Uozumi [49] applied the method to the pendulum equation with the assistance of the delay coordinate embedding techniques [69], proving that the time necessary for stabilization is reduced, whereas Yagasaki and Yamashita [94] obtained a faster stabilization of the considered UPO without using delay coordinate embedding techniques, even when some modeling errors or the influence of noise exist.

Wang and Jing [95] applied the Lyapunov function method to design a controller able to convert the chaotic motion of a pendulum to any periodic orbit in a shorter time than that required by OGY-based methods (much shorter than OGY but also shorter than SOGY) for the same parameter values and initial conditions. In particular, the Lyapunov method is able to overcome a difficulty arising when employing even modified versions of OGY method, which make use of the generalized Poincaré map, for stabilizing higher-periodic orbits, namely the unclearly specified stability directions – which play a basic role in the control procedure – around a multiple saddle solution.

Alasty and Salarieh [96] designed a non-linear feedback controller to stabilize arbitrary desired periodic orbits such as period-one, period-two, period-four orbits and more in a chaotic pendulum, and showed that the major advantage of this method is its shorter chaotic transient time compared with other methods such as OGY.

A variation of the discrete OGY technique called semi-continuous control (SCC) method, proposed by Hübinger et al. [85] and extended by DeKorte et al. [87], was considered by Pereira-Pinto et al. [97] to stabilize unstable periodic orbits – identified by the closed-return method – in a nonlinear pendulum with torsional stiffness and damping. In implementing the control, two different situations were considered, first, if either all state variables are available, as it happens when considering signals generated by numerical integration of a mathematical model based on experimentally identified parameters, or, second, just a scalar time series obtained from an experimental setup. In the second situation, state space reconstruction is used with the method of delay coordinates [69]. Of course, experimental data is associated with noise contamination which is unavoidable in cases of data acquisition, so that the effect of noise on control techniques is an important point to be analyzed in order to ascertain their robustness, which is an essential aspect to the controllability of a dynamical system. By analyzing the effect of noise on the controlling procedure the authors defined some relevant limitations. Nonetheless, the possibility of using this approach to control chaotic behavior in mechanical systems was confirmed.

Working on the same line, Pereira-Pinto et al. [98] proposed the use of "extended state observers" – which is a tool from control theory for determining non-observed states of a dynamical system – to perform the state space reconstruction from one scalar time series, instead of using the delay coordinate method. Again, they suc-

cessfully applied this technique in the framework of the semi-continuous control of chaos of the nonlinear pendulum.

3.2 Smooth Archetypal Oscillators

Since the very first formulation of the OGY method, a number of authors considered some kind of magneto-elastic beam as the archetypal experimental model for control of chaos in smooth mechanics.

The practicality of the method to lock a system response onto specific periodic orbits embedded in a chaotic attractor and move it from one periodic orbit to another was first demonstrated by Ditto et al. [99] for a chaotically oscillating magneto-elastic ribbon, whose effective Young's modulus is sensitive to small magnetic fields. By providing a time-dependent perturbation to the applied field (based on the position of the beam), Ditto et al. could stabilize the ribbon onto period-1 and period-2 orbits.

Referring to the classical experimental magneto-mechanical oscillator consisting of a mass attached to a cantilevered elastic beam [100], which exhibits a double-well potential and a two-dimensional Poincaré map, Moon et al. [101] used the "occasionally proportional feedback" technique developed by Hunt [102] as an extension of the OGY method to stabilize the system onto a period-one motion.

Also stimulated by that classical experiment, Dressler et al. [103] demonstrated the tracking of an UPO in a horizontally cantilevered, periodically driven, elastic bronze ribbon. Stabilization of the UPO at each tracking step is performed via the local control method, which is a variant of OGY. Starting with feedback control vectors extracted from the analysis of the experimental data at each tracking step, the location of the UPO is redetermined using an adaptive orbit correction that exploits the applied control signal and the actual trajectory of the system. In this way, the UPO is tracked over a broad parameter regime where the chaotic attractor has disappeared and another periodic orbit has become stable.

A physical system made of a magnetoelastic metal ribbon was also considered by Ding et al. [104], along with a pair of coupled Duffing oscillators with periodic forcing, to demonstrate the viability in practical applications of a method of chaos control in high dimensions. Using small time-dependent perturbations of a single system parameter, the method achieves control by stabilizing a desired unstable periodic orbit with any number of unstable directions, which is a situation where the OGY method does not succeed. At the same time it is shown how, also in situations (e.g., just one stable and one unstable direction) where the low dimensional control is effective, a higher dimensional implementation may prove to be more efficient.

In turn, a number of numerically oriented works were devoted to controlling chaos in smooth nonlinear systems such as the escape and/or Duffing oscillators, via some variants of the OGY method.

Investigating the possibility to control chaos in a temporally irregular environment, Ding et al. [43, 105] considered the softening Duffing oscillator driven

by an irregular excitation as a model of ship rolling under lateral ocean waves. Heuristically incorporating a short-term monitoring and prediction feature of the future evolution of the environment into the OGY-SOGY control scheme, the authors were able to stabilize a period one orbit and prevent the ship capsizing that would occur, in absence of control, upon the boundary crisis of a chaotic attractor.

One interesting topic addressed in [54] is the comparison of the OGY method with respect to other control techniques, which is very useful because it permits one to judge the good/bad performances of the method, and the range of its applicability. More precisely, Sifakis and Elliott [54] considered four different control techniques applied to the classical Duffing oscillator: (i) open-loop periodic perturbation method, which consists in adding a periodic perturbation to the system excitation, (ii) continuous delayed feedback method, i.e., the Pyragas [56] method, (iii) the Hunt method [102, 106], and eventually (iv) the OGY method [9]. As to the first technique, in [54] the perturbation is chosen empirically by a trial-and-error procedure, whereas in other works it is optimally determined on the basis of a theoretical analysis relying on system dynamical properties [37].

3.3 Vibro-Impact and Friction Systems

Impact is often undesired in engineering systems, for it causes excessive fatigue, noise, or even direct failure. As a consequence, systems are designed to avoid operating conditions that are conducive to impact, where possible. Of course, the proper choice of system parameters at the design stage represents the most basic form of passive control. However, in many cases, the engineer is not given the opportunity to completely define and specify all system characteristics before manufacture. Unforeseen changes in the operating environment, damage from external forces, or modified system requirements can lead to unexpected and unwanted responses, such as a multiple impacting chaotic motion, to be possibly suppressed with a control technique. In fact, the reduction of impacting events is a potentially important practical benefit of using chaos control in the mechanical context [107].

Kalagnanam [108] applied the OGY algorithm to a spring-mass system that impacts with a sinusoidally vibrating table. Besides stabilizing its chaotic attractor on period-1 and period-2 orbits using small perturbations of the driving frequency, he demonstrated the ability to switch the chaotic system between the two orbits by controlled time dependent perturbations. Moreover, since the system exhibits long chaotic transients before settling onto the stabilized orbit, he exploited the exponential sensitivity of a chaotic system to small perturbations to control the duration of such a transient via the SOGY approach [46], which uses chaos to direct trajectories to targets and drastically reduces the average length of the chaotic transient.

Begley and Virgin [80] showed how, using the adjustable friction force as the control parameter, the linear map approximation of the OGY method may work in

a two-sided impact-friction oscillator for stabilizing either an ideally non-impacting response or a single-impact response corresponding to an UPO embedded in the chaotic attractor. Actually, while showing that, in mostly ideal cases, very little additional control effort (corresponding to low-energy effective control) is required for stabilization by the OGY method once the system locks onto the unstable orbit, they also discussed how in different practical situations the control of many other UPOs may be unattainable: this happens, for example, if a significant experimental noise is such as to push the system out of the linear region after control is established, or if large unstable eigenvalues, which are in fact common in impact systems due to substantial stretching of the phase space at impact, do occur.

One important point for vibro-impact systems is that the temporal evolution of the dynamical variables consists of smooth motions governed by a linear differential equation interrupted by a series of non-smooth impacts, with the trajectories being thus discontinuous in phase-space. Due to discontinuities, the hard part of the OGY control process, which applies only tiny perturbations on an available control parameter, is how to determine the value of parameter perturbation. For that, using the analytical solution of the differential equation and the impact rule, a transcendental (impact or discontinuity) map has to be determined, with the dynamical variables being computed at the impact instants. Referring to a fundamental discontinuity map with a square root singularity in the Jacobian, the so-called Nordmark map [109] that captures universal properties of impact oscillators near grazing, Casas and Grebogi [42] applied the OGY technique to select particular trajectories with a desired sequence of impacts and stabilized an unstable periodic orbit with one impact per period, involved in the grazing bifurcations, by adjusting slightly the parameter related to the external force. In turn, de Souza and Caldas [110], considering an impact oscillator or an impact-pair system with the amplitude of excitation as control parameter, used a transcendental map to specify the parameter perturbation needed to implement the OGY method for stabilizing an UPO embedded in the chaotic attractor.

Also for impact systems, a number of works were just explicitly "motivated" by OGY control or used some kind of "OGY-like" methods. Thus, Bishop et al. [111] developed a control strategy to stabilize period-1 impacting motions in an impacting driven beam undergoing complex motion associated with grazing bifurcations, as the driving frequency is varied. Galvanetto [112] adapted to non-smooth systems a controlling algorithm previously developed for smooth systems [92], and used it for finding UPOs of a stick-slip system made of a chain of blocks on a moving belt and for stabilizing them by small changes of one or more control parameters.

Motivated by the need to understand the nature of some nonlinear resonances in complex, multiactuated, servo-hydraulic structural engineering applications, Gutiérrez and Arrowsmith [113] analyzed the control of a mathematical model of the laboratory setup of a double-impacting system, studying the stability of low-order impact resonances in the context of the application of control techniques using displacement feedback. Preservation or annihilation of experimental and analytical resonant periodic orbits was realized via control schemes based on the OGY

method, showing how the dissipative nature of impact resonances makes it possible to apply a low-dimensional analysis.

It is also worth mentioning that vibro-impact systems with oscillating parts colliding with rigid walls or with other vibrating components may have important practical applications. This is the case of impact dampers, where the vibration of a primary system is controlled by the momentum transfer through collisions with a secondary loose mass which bounces back and forth [114]. Impact dampers are used to control high-amplitude oscillations, such as those typically occurring in chaotic motion, and hence can be regarded as devices for controlling chaos in mechanical engineering systems [115], as in cutting tools, turbine blades, and chimneys. No direct application of the OGY method seems to have been made in this respect, to the authors' knowledge. However, de Souza et al. [116] studied the control of chaotic impacts of a vibrating cart driven by a non-ideal motor, obtained by adding an impact damper consisting of a bouncing particle. The transfer of momentum which follows each collision between the particle and the walls attached to the cart is responsible for an effective coupling so that the motion of the cart can be controlled by the bouncing motion of the particle. With a very small value (about 0.5%) of the ratio between the masses of the bouncing particle and the cart, chaotic attractors are replaced by two stable trajectories with a kind of stabilization "resembling that achieved by the OGY technique". Of course, from the technical point of view, the smaller the ratio of the masses, the easier it is to install such impact dampers in engineering systems like turbine blades and cutting tools. The authors also pointed out how, with respect to previous studies [114], the value of this ratio is considerably lowered due to the non-ideal character of the forcing, which is a consequence of its limited energy supply.

Dry friction in mechanical oscillators can also lead to chaotic dynamics under periodic forcing. Moon et al. [117] demonstrated control and anti-control of chaos, namely stabilizing an UPO in a strange attractor or driving the system into a chaotic state near a periodic motion, in an experimental dry-friction oscillator obtained by adding lateral titanium friction plates to the cantilevered elastic beam with attached mass considered for control in Moon et al. [101]. Anti-control is often beneficial in engineering devices such as robotic manipulators or cutting tools for normal machining of metals, where it corresponds to adding small noise or dither onto periodic oscillations, e.g., to break static friction or to avoid hysteretic behavior. Control and anti-control were based again on the "occasionally proportional feedback" technique [102], and were effected by changing the normal force of the dry friction element via a magnetic actuator.

3.4 Coupled Mechanical Systems

A number of coupled systems have been studied for control by using OGY-like methods.

Considering a kicked double rotor system made of two connected massless rods, Feudel et al. [118] exploited its ability to access many different coexisting states,

combined with its sensitivity and flexibility, to gear the dynamics toward a specific periodic behaviour. They used a simple feedback control scheme making use of perturbations smaller than those of the standard OGY, and also smaller than the noise amplitude.

In a tutorial paper devoted to controlling chaos in a dynamical system constituted by two coupled dynamos, Agiza [119] also dealt with suppression of chaotic behavior to one of the UPOs embedded within the attractor but this was realized by nonfeedback or a delay feedback control method which does not require finding the UPO to be stabilized but needs its period only.

Control and stabilization of an UPO of a chaotic system using a fuzzy system constructed on OGY method were investigated in Alasty and Salarieh [96], who applied the method to a Bonhoeffer-van der Pol oscillator representing a model of a two degree-of-freedom system kept in a fluid flow. Attention is focused to the advantages to be possibly obtained via fuzzification rather than on the application. It turns out that the transient response time and the control effort for stabilizing the period-4 unstable orbit with fuzzy-OGY controller are much less than with pure OGY control. This is due to the fact that when using fuzzy control the perturbation signals applied to the system are not sensible to the distance from the fixed point of the periodic trajectory in phase space. Fuzzification of OGY perturbing signal causes a smoother perturbing signal in larger domains of time and Poincaré section. The main role of a fuzzy system is to make a control perturbing signal in the points which are far from the fixed point, while in the near vicinity of fixed point the perturbation made by the fuzzy system is similar to that made by the pure OGY controller. Robustness of controlled system against random disturbances also increases with fuzzy-OGY controller. Comparable favorable features are exhibited, to a certain extent, by a fuzzy-Pyragas controller with respect to the pure Pyragas one.

A number of mechanical systems are actually high-dimensional, as it happens, e.g., in robotics. Underactuated manipulators arise in a number of important applications such as free-flying space robots, hyper-redundant manipulators, snake-like robots, and manipulators with structural flexibility, among others. An important goal in manipulator control is to perform tasks involving the exact tracking of some desired trajectory. After a proper analysis of the particular device dynamics, exact tracking depends on the nature of the designed control algorithm. The absence of an actuator transforms the robot into an underactuated device that may lead to a malfunction of the system, to an erroneous tracking of the desired trajectory and in some cases to instability. Moreover, underactuated robots may exhibit rich dynamical behaviors including chaos [120, 121].

In addressing these issues, González-Hernández et al. [120, 121] dwelled on some limitations of earlier control techniques for stabilization of periodic orbits, including the OGY one, which are designed for systems described by three differential equations yielding two-dimensional Poincaré or Lorenz maps and are only capable of imposing a particular dynamics for stabilization. Overall, they noticed how a general framework for control of higher-dimensional systems is not yet established.

In their paper, the authors aim at controlling higher-dimensional systems and at imposing a desired dynamics for stabilization. They report on the use of the Lorenz, instead of the Poincaré, map for the local identification of the dynamics around the UPOs in a chaotic attractor, claiming that its use allows for better identification schemes and avoids ill-conditioned identification problems [122]. As a first stage for achieving stabilization of periodic orbits embedded in higher-dimension chaotic attractors, based on system measurements, they propose a framework for designing flexible control laws which still belongs to the class of parameter perturbation methods. Flexibility means the capability to impose arbitrary dynamics with the proposed control law, which is applied to the so called Pendubot, namely an underactuated robotic system with two links. The proposed method offers a way to control chaotic systems without any prior knowledge of their dynamics, and needs a measure of only one of the system variables and the availability of a system parameter which can play the role of a control input.

3.5 Targeting in Astrodynamics

Besides strict suppression of chaos, one important outcome of the OGY approach to chaos control has been the flexible and efficient exploitation of the complexity inherent in the chaotic dynamics for their technological uses. In particular, the inherent exponential sensitivity of chaotic time evolutions to tiny perturbations has been exploited for targeting, i.e. for rapidly directing a system to a desired accessible state [46]. Major relevant achievements have been obtained in the field of astrodynamics, particularly related to the important and timely issue of spacecraft transfer, to be realized with a low consumption of propellant or energy.

In general, controlling and targeting in chaotic Hamiltonian systems – which are the situations faced in astrodynamical problems related to spacecraft transfer – are not easy to accomplish. Besides the coexistence of interwoven chaotic and quasi-periodic regions, the phase space is divided into layered components, which are separated from each other by Cantori [123]. Typically, a trajectory initialized in one layer of the chaotic region wanders in that layer for a long period of time before it crosses the Cantori to wander in the next layer. Several targeting strategies have been proposed to overcome these difficulties, e.g., using small perturbations [124] to drive a trajectory from an unstable periodic orbit located inside a chaotic scattering region to a target point outside it, which is a major problem [125], and then applying in the "controllable region" a method of control of chaos derived from OGY [126]. Macau [127] applied the overall procedure to a Hamiltonian system in a "soft" chaotic Hamiltonian evolution, i.e., targeting in the planar, circular, restricted three-body problem which models the dynamics of a spacecraft moving in the Earth-Moon system and is a special case of the full three-body problem, where one of the masses is taken to be infinitesimal and has no influence on the two primaries which are on circular orbits.

A review of the relevancy and efficiency of the OGY strategy of chaos control to spacecraft steering in different Hamiltonian situations is presented in Macau and

Grebogi [44], who also reported on the Hill problem governing the encounter phenomenon of two light bodies describing circular orbits around a heavy central body. In this case, the unstable periodic orbits embedded in a non-attracting chaotic invariant set related to chaotic scattering are exploited for in-orbit stabilization, namely to keep one satellite in orbit around the other [128].

It is worth stressing the importance and efficiency of control of chaos and targeting techniques as regards low-energy spacecraft transfer and in-orbit stabilization. Indeed, since the lifetime of an automatic space exploration mission is mainly defined by the amount of fuel it can carry, the direct benefits of chaotic transfer in astrodynamics are twofold, i.e., either allowing a spacecraft to go farther in exploring the limits of our Solar System, or saving the amount of fuel, which entails a larger amount of scientific instrumentation in the payload, since the liftoff capacity of a launch vehicle is limited. These techniques can also be used to rescue spacecrafts that end up in a wrong trajectory accidentally, due to defects in the launch phase.

3.6 Atomic Force Microscopy

Atomic force microscopy (AFM) is widely used for nano-scale material characterization and surface inspection in engineering applications. Indeed, as an imaging tool, AFM is capable of resolving surface features at the atomic level for conducting and non-conducting samples, and it is currently used in many imaging applications ranging from biological systems to semiconductor manufacturing.

The mechanism of AFM basically depends on the interaction of a micro-cantilever with surface forces. The tip of the micro-cantilever interacts with the surface through a surface-tip interaction potential. One approach to measure the surface forces is to monitor the deflection of the micro-cantilever through a photodiode. This approach is named “contact mode”. Another approach termed “tapping mode” is performed by vibrating the micro-cantilever close to its resonance frequency, at a constant driving amplitude, and monitoring the changes in its effective spring constant.

As reported in the introduction of Arjmand et al. [129], a micro-cantilever in tapping mode may exhibit chaotic behaviour in some regions of its physical parameters (damping value, excitation amplitude and frequency, average tip-sample distance), as observed experimentally [130] and verified theoretically [131] and numerically [132]. Yet, chaotic behavior is highly undesirable for AFM performance because it entails inaccurate measurements and low resolution of the achieved sample topography. Accordingly, it is always required to eliminate the possibility of chaotic motion of the micro-cantilever either by changing the AFM operating conditions to a region of parameter space where regular motion is assured or by designing an active controller that stabilizes the system on one of its UPOs. Arjmand et al. [129] developed a nonlinear delayed feedback control algorithm to stabilize the AFM micro-cantilever on its first-order UPO, as well as a delayed feedback control via sliding mode scheme for chaos elimination.

In turn, Misra et al. [133] documented a variety of dynamical systems-based, control-theoretic analytical, computational and experimental tools for exploiting the natural oscillating dynamics of the cantilever in tapping-mode operation of AFMs. In particular, they applied different OGY-based feedback strategies to the AFM cantilever dynamics described by a lumped-parameter model in which the spatial shape of the vibrating beam is approximated by its first fundamental mode of vibration. By imposing discrete changes in the vertical offset between the cantilever support and the sample surface based on an estimated linearization of the system dynamics about a dynamically generated reference trajectory, they aimed at maintaining a desired oscillation amplitude of the cantilever and, indirectly, at extracting information during surface scanning of the sample profile. Three different control schemes, full-state feedback, estimated-state feedback and partial-state feedback, are examined in terms of their capability of stabilizing unstable periodic oscillations, reducing the extent of transient dynamics in the vicinity of stable periodic oscillations, and maintaining a desired periodic oscillation while scanning the sample surface at fairly high lateral speeds, under the influence of a train of disturbance inputs in the form of surface profile variations. Numerical results document increases in the speed of surface scanning, while preventing undesired transitions between different cantilever oscillations, as well as successful use of (in the absence of control) unstable cantilever oscillations that achieve tapping-mode operation with relatively small contact velocities and short duration of contact.

4 An Alternative Unified Framework for Control of Non-regular Dynamics of Mechanical Systems

One characteristic of the OGY method is that it *exploits* the chaotic properties of systems to achieve some desired performances, which in the specific case is the stabilization of an unstable saddle embedded in the chaotic attractor.

There are several other methods which likewise exploit the chaotic performances of dynamical systems to obtain certain goals, although they refer to different features with respect to the OGY method. Among them, an important role is played by the methods aimed at eliminating, or shifting in parameters space, a given homoclinic or heteroclinic intersection which triggers unwanted dynamical phenomena [17, 37, 41, 134].

The objective of eliminating the homoclinic bifurcation can be reached in various manners, namely,

1. by adding a controlling parametric excitation to an uncontrollable external excitation, or viceversa [41];
2. by modifying the excitation by adding controlling terms to a given uncontrollable one. This can be done, for example, by adding single or multiple subharmonics or superharmonics (or even ultra-subharmonics) to a given harmonic term [17, 37], and even by choosing these added terms in an optimal way [37];
3. by modifying the parameters of the system.

The last case is trivial from a theoretical point of view, and it is mentioned only because it can have some interest from a practical point of view. It just requires one to identify the chaotic region in the parameters space, and to change the parameters to stay outside that region.

The first two cases, on the other hand, are not very different from a conceptual point of view. The differences are mainly (but not uniquely) technical: the common idea is that of adding controlling terms to eliminate a given, unwanted, homo/heteroclinic intersection. So, we can describe just one of them, and we will illustrate the technique developed by Lenci and Rega in a series of papers [26, 37, 62, 63, 107, 135–146].

Any kind of control method based on the elimination of homoclinic or heteroclinic intersections, or better on shifting or eliminating the global bifurcations through which complex dynamics appear, requires various steps, both of theoretical and practical nature. Although they can be supported by analytical arguments, herein they are mostly discussed at a phenomenological level with the aim of introducing the general ideas and features of the method, usually not reported in regular publications, and of highlighting its overall framework. Of course, the various technical aspects can be found in the specific papers.

4.1 Single Degree-of-Freedom Systems

With the aim of capturing the key features without redundancies we refer to single degree-of-freedom (d.o.f.) models, which describe the dynamics of simple mechanical systems or of reduced order models of complex structures.

This choice may appear very restrictive at a first glance, but it is not. In fact, in general, the additional structure due to the higher dimensionality does not add any significant element to the control *ideas* and to the theoretical apparatus, but contributes mainly *technical* difficulties, which we want to by-pass because they are not due to the control, or can possibly add new resources, thus enlarging instead of reducing the applicability of control.

For example, in large dimensions, it may be difficult to practically measure the distance between stable and unstable manifolds, which is necessary to detect the considered bifurcation. This implies that it may be difficult to *practically* apply control to those structures, but this does not affect the *theoretical* characteristics of control, and the possibility of its application. What can instead occur is to take advantage of the complexity of the structure to look for further properties of the control, which add to those investigated with single d.o.f. models.

According to the previous considerations, the generic system which will be considered is governed by the ordinary differential equation

$$\ddot{x} + f_1(x) + f_2(x,\dot{x}) + f_3(x,t) = 0, \tag{1}$$

whose solution $x(t)$ represents the status of the system. Four terms, which are the main "components" of a common mechanical system, occur in Eq. (1): inertia, $\ddot{x}$,

conservative elastic restoring force, $f_1(x)$, dissipation or damping, $f_2(x,\dot{x})$, and external or parametric excitation, $f_3(x,t)$. Note that in general the term $f_2(x,\dot{x})$ may also provide a negative damping, i.e. it can pump energy into the system. However, in this work, we will consider only standard mechanical systems where $f_2(x,\dot{x})$ dissipates energy.

When f_2 and f_3 vanish, the system (1) is Hamiltonian, or "conservative". In this case, the Hamilton function $H(x,\dot{x}) = \dot{x}^2/2 + \int f_1(x)dx$ is constant along orbits. In mechanical language H is the total energy, sum of the kinetic $K(\dot{x}) = \dot{x}^2/2$ and potential $V(x) = \int f_1(x)dx$ energy. The terms f_2 and f_3 represent two different kinds of modifications, or perturbations, of an Hamiltonian system: those which extract energy from the system, and those which pump energy into the system, respectively.

More difficult situations can of course be found in the reality, but the case (1) is sufficient for our purposes.

4.2 Different Kinds of Global Bifurcations

The first consideration, less obvious than what might be thought, concerns the relevance of the global bifurcations for the complex behavior. In fact, we can have two different situations: bifurcations not directly related to a specific outcome in the system dynamics, and bifurcations immediately causing a strong variation in the response.

While actually being the same dynamical phenomenon, there are strong differences between the two classes of bifurcations. The former usually triggers chaotic mechanisms, such as, for example, the creation of a chaotic saddle, which do not immediately, and dramatically, influence the response, but which constitute the necessary pre-requisites for successive routes to chaos. Their effects are hidden behind the visible outcome, but they are not negligible, even from a practical point of view.

Remarkable examples of this case are the homoclinic or heteroclinic orbits of hilltop saddles of Hamiltonian systems. Their invariant manifolds usually surround potential wells where the important part of the dynamics occurs, and constitute a barrier against the penetration of trajectories from out-of-well attractors eroding the well. Once this barrier has been broken by the global bifurcation, the (chaotic) effects can penetrate the well and, sooner or later, become visible. From the one side, this permits deeper understanding of the dynamical mechanisms involved in the event, especially with respect to safe well erosion [146, 147], and from the other side it permits detailed and specific investigations, to be followed by smart application of the control method.

Due to the underlying Hamiltonian nature, these bifurcations can be usually detected analytically by the Melnikov method [48] or even directly in closed form. This is important because it permits a more general application of the control method [62, 137].

The second class of bifurcations is involved in specific boundary or interior crises, and determines sudden appearance/disappearance or enlargement of chaotic

attractors or sudden modifications of their basins of attraction. Thus, their elimination directly entails elimination of the crises, with immediate, visible, effects [136]. These are the situations where the class of control methods discussed in this section are very effective. Here in fact even a small control leads to remarkable results, and, e.g., the chaotic attractor can be easily removed/created.

The drawback consists in the fact that these bifurcations are rarely related to hilltop saddles, and commonly cannot be easily detected or even discovered. In fact, it can (and actually does) occur that the saddle S directly responsible of a crisis is hidden by other, possibly more robust, saddles with an heteroclinic connection with S. This can generate an heteroclinic tangle [148], which makes any attempt to detect the saddle and its invariant manifolds a challenge from a practical (i.e. numerical) point of view, and even a somehow questionable theoretical issue.

In the most fortunate cases where, in spite of previous difficulties, the detection is possible, it usually requires long, time consuming, and systematic numerical investigations, which are cumbersome and do not permit a full understanding of control or its full exploiting. Thus, the application of control is less general than in the former case.

However, it should be clear that, strictly speaking, the detection of the saddles and of their manifolds, which is a necessary pre-requisite for applying control, is not a part of the control method, which can actually be developed, in its general lines, without any reference to a specific situation (and even, in some cases, without references to any specific systems). The ability of detecting the global bifurcations directly responsible for the observed crises simply provides a more powerful possibility of eliminating/creating chaos by a straightforward application of a control method.

The previous considerations explain why we will deal with homoclinic or heteroclinic bifurcations of hilltop saddles (but see also [136] where the control is applied to a numerically detected homoclinic bifurcation of a non-hilltop saddle). In fact, although they are involved in boundary or interior crises only in some situations, they are relatively easy to be analytically detected by the Melnikov or similar method, thus permitting a detailed development of the most important aspects of control, without reference to any specific situation. To support this choice, we further underline that the considered global bifurcations trigger the dynamical events eventually leading to chaos, so that its elimination is useful in any case.

4.3 Distance Between Stable and Unstable Manifolds

From a practical point of view, the key to applying any kind of control based on the elimination of homoclinic or heteroclinic bifurcations is the detection of distance between the stable and unstable manifolds.

To illustrate the main properties, we refer to an homoclinic bifurcation of a nearly Hamiltonian system (Fig. 1), i.e., a conservative system with small perturbations, which is supposed to have a saddle point with an homoclinic orbit. The major aim is that of verifying whether the homoclinic orbit survives or not when damping and

excitation, here seen as perturbations, are added. The objective of this section is indeed determining the distance between perturbed stable and unstable manifolds.

In the absence of perturbations, the stable and unstable manifolds of the saddle coincide because they are constrained to have the same energy (Fig. 1a).

4.3.1 Effect of Damping

When damping is added to the Hamiltonian system, the coinciding stable and unstable manifolds – which still exist – split away from each other (Fig. 1b). In fact, the saddle does not change energy (being a fixed point), the unstable manifold loses energy in forward time (and asymptotically approaches an attractor different from the saddle), and the stable manifold increases its energy backward in time (as it must approach the saddle as $t \to +\infty$ and its energy decreases forward in time). Thus, a homoclinic loop is no longer energetically possible and, on appropriate Poincaré maps, usually the stroboscopic one, there is a non zero distance between stable and unstable manifolds. It is easy to realize that this distance is proportional to the damping amplitude, herein denoted by $\varepsilon\delta$, and that in the case of small damping, it is approximately constant along a sufficiently short interval (Fig. 1b).

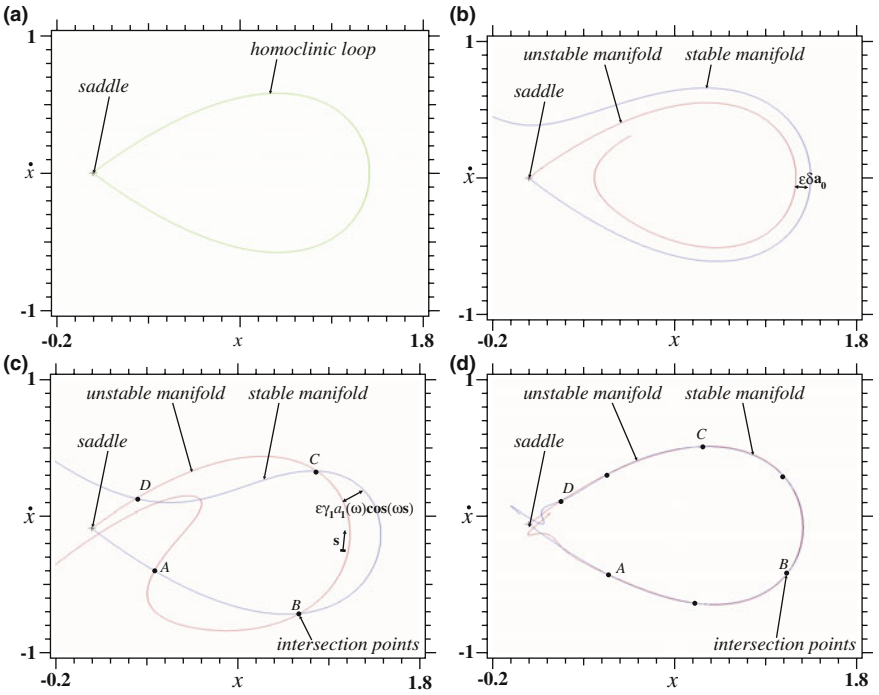


Fig. 1 Schematic illustration of the distances between stable and unstable manifolds in an homoclinic bifurcation of a nearly Hamiltonian system. (**a**) Hamiltonian case; (**b**) damped case; (**c**) harmonically excited case, excitation frequency = ω; (**d**) superharmonically excited case, excitation frequency = 2ω

4.3.2 Effect of Excitation

The second standard perturbation is the excitation $f_3(x,t)$, which is assumed to be periodic in time with angular frequency ω and zero mean value (a constant excitation simply shifts the reference/equilibrium point of the system). Furthermore, it is initially supposed to be harmonic and state independent or "external," namely, $f_3(x,t) = \varepsilon\gamma_1 \sin(\omega t)$, $\varepsilon\gamma_1$ measuring the excitation amplitude. The extension to state dependent "parametric" excitations can be obtained by referring to the so called *equivalent external* excitation in the evaluation of the manifolds distance, as shown in [137].

Both the definitions of the amplitudes $\varepsilon\delta$ and $\varepsilon\gamma_1$ are left somehow vague up to now, but the physical intuition is sufficient for the purposes of the present section. We only stress that they are small by the introduction of the dimensionless smallness parameter ε.

The excitation affects the invariant manifolds by producing an oscillating distance on the Poincaré map, as shown in Fig. 1c, where some of the primary intersection points (P.I.P., see [148]) are marked by A, B, C and D. Note that, according to common sense, the wavy pattern of the excitation produces a wavy pattern of the distance, and this is confirmed by Fig. 1d, which shows that doubling the frequency of the excitation doubles the number of P.I.P.

For what concerns the magnitude of the distance, it is clear that it is directly proportional to the excitation amplitude, and that for small amplitudes it depends linearly on it. Moreover, comparison of Figs. 1c and 1d shows that the magnitude of the distance also depends on ω, and in the particular case of Fig. 1 it (rapidly) decreases with increasing ω. In other cases it can instead increase with ω, but anyway *the magnitude of the distance strongly depends on the excitation frequency*.

4.3.3 The Perturbed Manifold Distance

Based on the previous considerations, and focusing attention on the interval $A-B-C-D$ of Fig. 1c, d, we can write the distance between stable and unstable manifolds in the following approximate way (see Fig. 1b, c):

$$d(\omega s) = \varepsilon\delta a_0 + \varepsilon\gamma_1 a_1(\omega) g(\omega s), \tag{2}$$

where s is an arc length parameter (Fig. 1c), a_0 and $a_1(\omega)$ are a constant and a function depending on the specific dynamical system, and ε stresses that the assumed linearity with respect to δ and γ_1 holds only for small damping and excitation amplitudes. The smallness is also tacitly assumed in the additive form of the distance. Finally, in Eq. (2) $g(\omega s)$ is the oscillating part of the distance. It is due to the periodic excitation, and like it, it has zero mean value and is periodic with the same frequency. Since the excitation is harmonic and with small amplitude, it is possible to approximate $g(\omega s)$ with an harmonic function, e.g., $g(\omega s) = \cos(\omega s)$. Using cosine instead of sine has an unessential technical reason, linked to the choice of the point for which $s = 0$ (Fig. 1).

Equation (2) captures two opposite effects suggested by the physical intuition, according to which damping gives a positive distance, thus contributing to manifolds non-intersection, whereas harmonic excitation gives an oscillating distance, thus contributing to manifolds intersection.

4.3.4 More General Excitation and Damping.

The extension of the previous consideration to the case of a non-harmonic, but still periodic and external, excitation

$$f_3(x,t) = \varepsilon \sum_{j=1}^{N} \gamma_j \sin(j\omega t + \Psi_j) = \\ = \varepsilon\gamma_1 \sum_{j=1}^{N} \left(\frac{\gamma_j}{\gamma_1}\right) \sin(j\omega t + \Psi_j) = \varepsilon\gamma_1 f(\omega t), \tag{3}$$

is straightforward. In fact, by using the linearity, consequence of the smallness of the amplitudes, we have that

$$d(\omega s) = \varepsilon\delta a_0 + \varepsilon \sum_{j=1}^{N} \gamma_j a_1(j\omega) \cos(j\omega s + \Psi_j) = \\ = \varepsilon\delta a_0 + \varepsilon\gamma_1 a_1(\omega) \sum_{j=1}^{N} \frac{\gamma_j a_1(j\omega)}{\gamma_1 a_1(\omega)} \cos(j\omega s + \Psi_j) = \\ = \varepsilon\delta a_0 + \varepsilon\gamma_1 a_1(\omega) h(\omega s), \tag{4}$$

More generally, the extension to arbitrary – but small – damping $f_2(x,\dot{x})$ and excitation $f_3(x,t)$ leads to an expression of the distance with the same structure as (4). The differences are only of technical nature, and appear in the expressions of a_0 and $a_1(\omega)$, and therefore in a different relation between the Fourier coefficients of the excitation and those of the distance between the manifolds. In fact, it is not difficult to accept that a generic periodic excitation gives an oscillating distance, and that the shape of this oscillation is due to the shape of the excitation in a way which can be computed, at least in principle. Therefore, from now on we will refer directly to (4) without specifying the nature of the system, of the damping and of the excitation, which indeed are "hidden" behind (4). A schematic representation of this distance is reported in Fig. 2. The constant part due to damping and the oscillating part due to the excitation are clearly shown.

Equations (3) and (4) are written in the last forms because we assign $\varepsilon\gamma_1$ the role of overall excitation amplitude, while the remaining functions $f(m)$ and $h(m)$ measure the shape of the excitation and its effects on the manifolds distance. Note

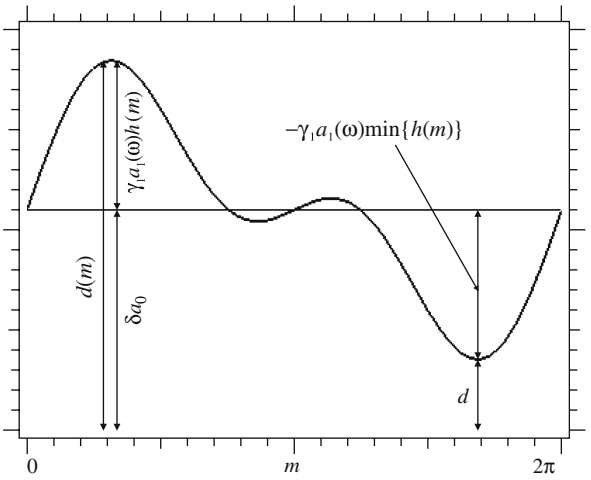


Fig. 2 The schematic illustration of the distance between stable and unstable manifolds

that, by definition, both $f(m)$ and $h(m)$ are 2π-periodic and $f_1 = h_1 = 1$, while $h_j = \frac{\gamma_j}{\gamma_1}\frac{a_1(j\omega)}{a_1(\omega)}$ (f_j and h_j are the Fourier coefficients of the respective functions). Furthermore, Ψ_1 represents an unessential phase shift, and can be assumed equal to zero without loss of generality (it is sufficient to consider $\hat{\Psi}_j = \Psi_j - j\Psi_1$ instead of Ψ_j).

4.3.5 Reference (Natural) and Controlling Excitations

The previous assumption of considering a general (although still periodic) excitation is important for the development of control, and is now briefly discussed. In fact, if one wants to operate by modifying the excitation, for example, by adding to a given natural excitation a properly designed controlling excitation, as done in classical control [60], one must have clearly in mind that the natural excitation cannot be touched, otherwise the problem becomes trivial. In other words, one must have a reference, unchangeable, excitation with respect to which to measure the improvements due to the control. On the contrary, it is trivial to change the excitation in order to get a desirable dynamical response: it is sufficient to look at the behavior chart of the system and chose the excitation (or its parameters) which provides the desired response. E.g., whatever the reference excitation is, if we are able to change its amplitude it is obvious that we can easily move from chaotic to non-chaotic response regimes.

In this chapter, where only periodic excitations are considered, the reference is the *harmonic* excitation, which is the most natural and intuitive. Thus, we implicitly think of the excitation as made of two parts,

$$f_3(x,t) = \varepsilon\gamma_1 \sin(\omega t) + \varepsilon\hat{f}_3(x,t) = \varepsilon\gamma_1\{\sin(\omega t) + \bar{f}_3(x,t)\}, \tag{5}$$

In general, the two parts have their own amplitudes, but we find it useful to refer to a unique excitation amplitude, as done in the last expressions of (3), (4) and now in (5), where the overall excitation is just that of the reference excitation. Other choices are possible in principle, but the present one seems to be the most natural in the field of the proposed control method.

Within the framework of periodic excitations, the choice of the harmonic one as the reference excitation is not restrictive. In fact, if another reference excitation is chosen, then Eq. (5) becomes

$$f_3(x,t) = \varepsilon\gamma_1\{f_{3,\mathrm{ref}}(x,t) + \bar{f}_{3,\mathrm{cont}}(x,t)\}, \tag{6}$$

where once more the amplitude of the reference excitation has been singled out and interpreted as the overall amplitude. By equating the two expressions of $f_3(x,t)$ given in (5) and (6) we get,

$$\bar{f}_{3,\mathrm{cont}}(x,t) = \sin(\omega t) - f_{3,\mathrm{ref}}(x,t) + \bar{f}_3(x,t), \tag{7}$$

so that if we know the control excitation $\bar{f}_3(x,t)$ in the case of harmonic reference excitation, then we know the control excitation $\bar{f}_{3,\mathrm{cont}}(x,t)$ corresponding to any reference excitation $f_{3,\mathrm{ref}}(x,t)$.

4.3.6 Energetic Derivation of Perturbed Manifolds Distance

The expression (4) is previously obtained by qualitative considerations. In spite of this, it provides exact information, because it gives the same manifolds distance as computed by the Melnikov method [148]. Furthermore, it can be derived by energy arguments as follows (see [149] for similar reasonings).

An (hypothetic) homoclinic orbit $x_h(t)$ is possible if and only if the sum of the total energy E_d dissipated by damping and of the total energy E_p pumped by the excitation is zero. These energies are given by, respectively,

$$\begin{aligned} E_d &= \int_{-\infty}^{+\infty} f_2[x_h(t),\dot{x}_h(t)]\dot{x}_h(t)dt, \\ E_p &= \int_{-\infty}^{+\infty} f_3[x_h(t),t]\dot{x}_h(t)dt. \end{aligned} \tag{8}$$

Let us define the "energetic distance" $d = E_d + E_p$, so that a condition for $x_h(t)$ to be an homoclinic orbit is

$$d = \int_{-\infty}^{+\infty} f_2[x_h(t),\dot{x}_h(t)]\dot{x}_h(t)dt + \int_{-\infty}^{+\infty} f_3[x_h(t),t]\dot{x}_h(t)dt = 0. \tag{9}$$

Remark. An alternative, but equivalent, way to obtain (9) is the following. Multiply the field equation (1) by $\dot{x}_h(t)$ and integrate in time from $t = -\infty$ to $t = +\infty$. We get $H(+\infty) - H(-\infty) + d = 0$, where H is the Hamilton function. Since the orbit is

homoclinic, it approaches the same saddle for $t = \pm\infty$, so that $H(+\infty) = H(-\infty)$ and $d = 0$ follows.

If $d \neq 0$ the guessed $x_h(t)$ cannot be an homoclinic orbit. Recalling [148] that an homoclinic intersection between stable and unstable manifolds implies the existence of an homoclinic orbit, it is possible to conclude that if we are able to show that no solutions of (1) exist satisfying (9) and approaching the same saddle at infinity, then the manifolds keep disjoint and there is no homoclinic intersection. In other words, there is a link between the manifolds distance and the energetic distance, as they simultaneously vanish in the case of homoclinic intersection. By varying a system parameter, an homoclinic bifurcation occurs when both become zero for the first time.

The Eq. (9) holds in general, and it is an useful check for a guessed homoclinic solution. We specialize it to the considered case of a nearly Hamiltonian system. The unperturbed (Hamiltonian) system has an homoclinic orbit $x_0(t-s)$ (see Fig. 1), where t is the real time and s an arbitrary phase shift introduced because the reference system is autonomous (see Eq. (1) with $f_2 = f_3 = 0$). Let the perturbations f_2 and f_3 be small, say of the order of ε ($f_2 \to \varepsilon f_2$ and $f_3 \to \varepsilon f_3$, see, e.g., Eq. (5)). Then, it is natural to assume that, under certain regularity conditions, any solution of the perturbed system is ε-close to the solution of the unperturbed one to which it is expected to converge when $\varepsilon \to 0$. Accordingly, we assume

$$x_h(t) = x_0(t - s) + \varepsilon x_1(t) + ..., \tag{10}$$

where $x_1(t)$ no longer contains the arbitrary phase because it is a solution of a non-autonomous system.

Inserting (10) in (9) we get

$$\begin{aligned} d(s) &= \varepsilon\{\int_{-\infty}^{+\infty} f_2[x_0(t-s), \dot{x}_0(t-s)]\,\dot{x}_0(t-s)dt \\ &+ \int_{-\infty}^{+\infty} f_3[x_0(t-s),t]\dot{x}_0(t-s)dt\} + ... \\ &= \varepsilon\{\int_{-\infty}^{+\infty} f_2[x_0(t), \dot{x}_0(t)]\dot{x}_0(t)dt \\ &+ \int_{-\infty}^{+\infty} f_3[x_0(t), t+s]\dot{x}_0(t)dt\} + ..., \end{aligned} \tag{11}$$

namely, $d(s) = \varepsilon\{$constant+periodic function of $s\}$+higher order ε-terms. It is important to stress that this expression requires only the knowledge of the unperturbed homoclinic orbit $x_0(t)$, and that the constant part of (11) is due to the damping $f_2(x,\dot{x})$, while the periodic part of (11) is clearly related to the periodicity in time of the excitation $f_3(x,t)$.

It is immediate to check that (11) and (4), which have been obtained under the same hypothesis of nearly Hamiltonian system, not only simultaneously vanish

in the case of homoclinic intersection, but they also have the same mathematical structure and the same functional dependence on the perturbations. It is then possible to conclude that the latter provides the searched energetic justification of the former.

On the basis of the previous considerations we can even obtain formulas for a_0, $a_1(\omega)$ and $h(\omega s)$, which are the key elements of (4), by a direct comparison. It is easy to check that they provide the *same* expression of the classical Melnikov method [148], so that they are exact in the asymptotic sense.

4.3.7 Minimum Manifolds Distance

Independent of its phenomenological derivation, its physical meaning or its exact calculation by the Melnikov or other methods, we have found that to look for homoclinic intersections we have to search the zeros of the distance (4).

First, we note that, as we are interested in the zeros of (4) and not in its sign, it is possible to assume $a_0 > 0$ without loss of generality. Then, the distance is an oscillating function around a constant positive value, as schematically shown in Fig. 2. From this picture and from (4) it is seen that the distance $d(\omega s)$ has zeros if and only if its minimum

$$\begin{aligned} d &= \min_{m\in[0,2\pi]}\{d(m)\} = \varepsilon\delta a_0 + \varepsilon\gamma_1 a_1(\omega) \min_{m\in[0,2\pi]}\{h(m)\} \\ &= \varepsilon\delta a_0 - \varepsilon\gamma_1 a_1(\omega) M \end{aligned} \tag{12}$$

is negative. When d is equal to zero for the first time by varying a governing parameter, for example the excitation amplitude, an homoclinic bifurcation occurs.
Remark. The expression (12) is useful in the case in which a_0 and $a_1(\omega)$ have the same sign. If, on the other hand, they have a different sign, then we must consider

$$\begin{aligned} d &= \min_{m\in[0,2\pi]}\{d(m)\} = \varepsilon\delta a_0 + \varepsilon\gamma_1 a_1(\omega) \max_{m\in[0,2\pi]}\{h(m)\} \\ &= \varepsilon\delta a_0 + \varepsilon\gamma_1 a_1(\omega) M \end{aligned} \tag{13}$$

instead of (12).

In the expression (12) a key role is played by the number

$$\begin{aligned} M = M\left(\frac{\gamma_j}{\gamma_1}, \Psi_j\right) &= -\min_{m\in[0,2\pi]}\{h(m)\} = \\ &= -\min_{m\in[0,2\pi]}\{\sum_{j=1}^{N}\left(\frac{\gamma_j}{\gamma_1}\right)\frac{a_1(j\omega)}{a_1(\omega)}\cos(jm+\Psi_j)\}. \end{aligned} \tag{14}$$

It is positive, because $h(m)$ has zero mean value, and it does not depend on the overall excitation amplitude $\varepsilon\gamma_1$, which is singled out in (12). Indeed, it is just this property which suggested the previous choice of the overall excitation amplitude.

While being independent of the excitation amplitude, M is instead strongly dependent on the shape of $h(m)$ and, consequently, on the *shape of the excitation*. The most important property of M is that it summarizes in a unique number this dependence, or, equivalently, it summarizes the contribution of the superharmonics added to the reference harmonic excitation (see Eq. (5)).

In the case of the harmonic excitation we have $h(m) = \cos(m)$, so that $M = 1$. This number therefore represents a reference value for M. In fact, we can distinguish between three families of periodic excitations: those for which $M = 1$, $M < 1$ and $M > 1$. The first class is equivalent to the harmonic excitation for what concerns the homoclinic bifurcations. The second class is useful in the problem of control, as we will see in the sequel, while the latter can be used in the case of anti-control [141].

4.4 Influence of the Parameters on the Manifolds Distance

The most important result of the previous section, which is summarized by the expressions (12), is that the minimum distance between the stable and unstable manifolds has been computed as a function of the governing parameters. Roughly speaking, we can distinguish between various families of parameters.

The first family is constituted by the mechanical system parameters such as mass, linear and nonlinear stiffnesses, etc. In (12) they are embedded in the definitions of a_0 and $a_1(\omega)$. For a given mechanical oscillator they are known, while they can be modified in the design process. Their influence cannot be discussed in general, because (i) it strongly depends on the mechanical characteristics of the system (hardening vs. softening, etc.), and (ii) it depends on how a_0 and $a_1(\omega)$ depend on the mass and stiffnesses, and this considerably varies from system to system, as it can be seen by comparing the expressions obtained for different oscillators [37].

The second family consists of the main parameters of the perturbations, such as damping and excitation amplitude and frequency, which are instead singled out in (12). Also these parameters can be considered as data, for a given problem, or can be modified during the design. In the considered case of small perturbations, damping and excitation amplitude influence *linearly* the minimum distance (12), and so their effects can be easily understood. In particular, note that when the damping is zero with non zero excitation, d is always negative, namely, there is always homoclinic intersection in almost Hamiltonian systems (Fig. 1c, d). On the contrary, with damping and without excitation d is always positive so that there is no homoclinic intersection at all (Fig. 1b).

More involved is the influence of the excitation frequency. In this respect, we can only infer that $\lim_{\omega\to 0} a_1(\omega) = 0$, because $\omega = 0$ corresponds to static excitation which does not pump energy into the system, and that $\lim_{\omega\to\infty} a_1(\omega) = 0$, because when the frequency is very high, only the time average of the excitation actually forces any given inertial system: but since the excitation is periodic, the average

is zero, so that the system is basically not forced, and the oscillating term in the distance disappears.

The last family we consider is constituted by the excitation shape parameters γ_j/γ_1 and Ψ_j, which are the key ingredients for the forthcoming analysis. Their influence on the minimum distance is summarized by the number M, which basically plays the role of an interface between the excitation shape and d. Not only are these parameters very important in the control method which is the object of this section (because they are allowed to vary for control design purposes), but they are also the less common because usually the excitation is considered as harmonic (sinusoidal). Thus, they deserve special attention.

Actually, the dependence of d on M is linear, and thus it is trivial. The non trivial point is instead the (nonlinear) dependence of M on γ_j/γ_1 and Ψ_j. The things are quite involved because the dependence of M on γ_j/γ_1 is modulated by the factors $\frac{a_1(j\omega)}{a_1(\omega)}$ appearing in the definition of h_j. Since these factors are system dependent, because so is $a_1(s)$, we cannot discuss in general the dependence of M on γ_j/γ_1.

What we can do in a *unified* way, i.e. without referring to a specific mechanical system, is to analyze the dependence of M on h_j (and on Ψ_j). To illustrate this fact by an example, we consider $N = 2$, i.e. a single added superharmonic. The function $M(h_2,\Psi_2) = -\min_{m\in[0,2\pi]}\{\cos(m) + h_2\cos(2m + \Psi_2)\}$ is plotted in Fig. 3, from which we see that both h_2 and Ψ_2 strongly influence M. More precisely, we clearly see that there are regions for which $M < 1$, which will be useful for control, and regions for which $M > 1$, which are useless (dangerous indeed) with respect to the control features. These conclusions are general and hold also for the case $N > 2$.

4.5 Homoclinic Bifurcation Thresholds

The multidimensional parameters space can be divided in two regions, according to whether the manifolds minimum distance is positive, corresponding to non intersecting, or detached, manifolds, or the minimum distance is negative, corresponding to intersection. The boundary between these two regions, i.e., the locus of points $d = 0$, corresponds to manifolds tangency, namely, to the homoclinic, or heteroclinic, bifurcation. In fact, crossing this boundary the manifolds become intersecting (or detach, according to the direction of crossing).

The detection of this boundary is crucial for the control method based on the elimination of the homoclinic or heteroclinic bifurcations. In fact, the method can also be interpreted as aimed at shifting this boundary in a certain direction, usually toward higher excitation amplitudes.

It is common to represent the bifurcation threshold in terms of excitation amplitude, namely, to solve $d = 0$ with respect to γ_1. As a matter of fact, this is one of the reasons why we pay a lot of attention to the definition of a proper excitation amplitude along this section.

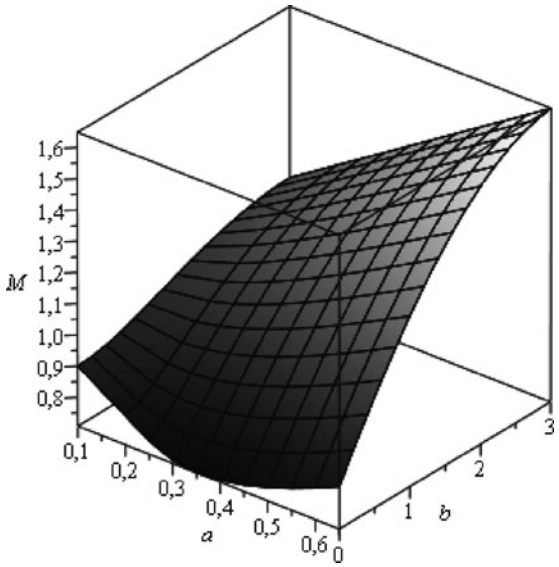


Fig. 3 The function $M(a,b)$, where $a = h_2 = \left(\frac{\gamma_2}{\gamma_1}\right)\frac{a_1(2\omega)}{a_1(\omega)} \in [0,0.65]$ and $b = \Psi_2 \in [0,\pi]$

The solution of $d = 0$ represents the critical excitation amplitude threshold for which the homoclinic or heteroclinic bifurcation takes place, and it will be denoted by $\gamma_{1,cr}$. If the distance can be expressed by (12) the computation of the critical threshold is trivial:

$$\gamma_{1,cr} = \frac{\delta a_0}{a_1(\omega)}\frac{1}{M}. \tag{15}$$

The simplicity of the previous expression is obviously a consequence of the linear nature of the distance with respect to γ_1, which however no longer holds in more general cases, such as, for example, those in which the manifolds distance is computed numerically [136]. Here, the solution of $d = 0$, while being conceptually simple, requires an extra computational effort consisting of solving a nonlinear algebraic equation. Furthermore, the equation is known only in a discrete set of points, and this is another element of difficulty.

In any case, irrespective of how the critical excitation amplitude has been computed, exactly, like in piece-wise linear systems [62, 140], by a perturbative approach, like in (15), or numerically, this value is eventually a function of all the other parameters of the system.

It is interesting that the excitation shape enters the expression (15) only by means of the single number M, which evidently summarizes also in this respect the effects of the superharmonics added to the basic harmonic excitation. Of course, this property is not expected to hold in general, and in other cases the dependence of $\gamma_{1,\mathrm{cr}}$ upon the excitation shape is more involved.

We remark that both the minimum distance d (through Eq. (12)) and the homoclinic bifurcation threshold $\gamma_{1,\mathrm{cr}}$ (through Eq. (15)) are inversely proportional to the excitation shape parameter M, i.e., the smaller M, the larger d and $\gamma_{1,\mathrm{cr}}$. The influence of the other parameters is complex, and in general it is far from linear. In fact, only the damping δ enters linearly in (15), this being a consequence of its smallness.

In general, not even a qualitative estimate is possible a priori. For example, for some mechanical systems the critical threshold of the homoclinic bifurcation of the hilltop saddle tends to a finite value when the excitation frequency tends to zero (e.g., the inverted pendulum between lateral barriers [62]) while in others it is unbounded (e.g., the Helmholtz oscillator [137]). This is a consequence, or perhaps a source, of the fact that chaotic behaviour – which is strongly related to global bifurcations – is not monotonic with respect to any parameter.

A special situation occurs for the class of excitations having $M = 1$, which includes the harmonic excitation. Since it is considered as the reference one, we indicate by $\gamma^h_{1,\mathrm{cr}} = \frac{\delta a_0}{a_1(\omega)}$ the corresponding critical threshold. According to this definition, Eq. (15) can be written in the equivalent form

$$\gamma_{1,\mathrm{cr}} = \gamma^h_{1,\mathrm{cr}} \frac{1}{M}. \tag{16}$$

This expression has the advantage of splitting into two independent parts the main dependences of $\gamma_{1,\mathrm{cr}}$ upon the system parameters. In fact, $\gamma^h_{1,\mathrm{cr}}$ contains all information regarding the mechanical system, such as mass, stiffness, etc., and the *reference* excitation, such as its kind (i.e., external, parametric, etc.) and frequency, while M contains information only on the excitation shape (or on the *controlling* part of the excitation).

This property does not hold in more general situations, for example when the critical threshold is computed by numerical methods without any assumptions on the smallness of the excitation amplitude. Here all the parameters of the systems are strongly coupled, and it is not possible to single out the dependence on the excitation shape.

When the excitation is periodic, irrespective of being harmonic (i.e., sinusoidal) or not, external or parametric, etc., it is usual to draw the critical amplitude as a function of the excitation frequency, for fixed values of all the other parameters. An example referring to the Helmholtz oscillator [137] is reported in Fig. 4. Here the excitation is periodic, and accordingly, the critical threshold is $\gamma^h_{1,\mathrm{cr}}$.

From Fig. 4 it is clear how the parameters subspace is divided in the two regions of homoclinic intersection and homoclinic detachment. The first one is the region of the so-called "Melnikov chaos," where various manifestations of chaos are expected: some of them, such as fractal basin boundaries, sensitivity to initial conditions, etc., will certainly occur, although to a different extent, while others, such as the chaotic attractors, are only theoretically feasible, but not necessarily exist in the whole region. When chaos – in all its manifestations – is considered as an undesirable

event, this region must be regarded as unfeasible, or dangerous, and care is needed in staying sufficiently far away from it.

The region of Melnikov chaos is "unbounded" and convex in Fig. 4. There are no reasons to believe that these properties are general, and both bounded and/or non convex sets of homoclinic intersection are expected to occur, at least in principle. Even non-connected sets are possible.

4.6 Control Ideas

So far we have obtained all the necessary prerequisites for application of the control method, i.e., we have related the manifolds minimum distance d and the homoclinic (or heteroclinic) bifurcation threshold $\gamma_{1,\mathrm{cr}}$ to the system parameters, and we are now in a position to describe the control ideas.

There are two conceptually different approaches, which however in some cases are equivalent and yield the same control implementation.

In a *fixed* situation, namely, for a given system, given operational conditions and, therefore, given system parameters, one can be interested in eliminating the stable and unstable manifolds intersection by the action of control. This entails acting on the system in such a way that the minimum manifolds distance passes from negative to positive. Global bifurcations are not directly involved in this case, although assuming that there is a continuous passage from the uncontrolled (intersection, negative distance) to the controlled (no intersection, positive distance) system, obvi-

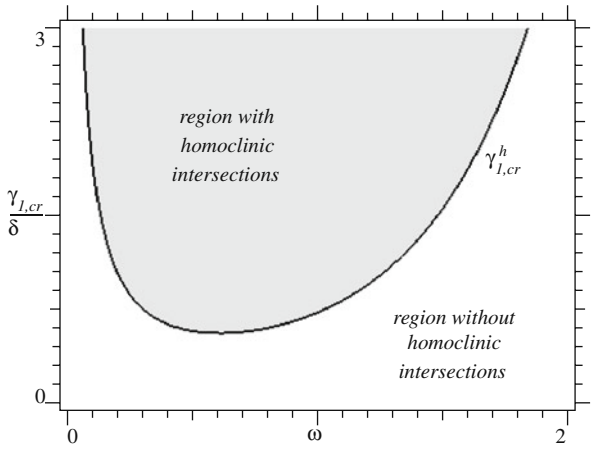


Fig. 4 The critical threshold $\gamma^{h}_{1,\mathrm{cr}}$ as a function of the excitation frequency ω, for the Helmholtz oscillator [137]

ously there is an intermediate condition in which the distance is zero, corresponding to a homoclinic (or heteroclinic) bifurcation.

This approach can be of interest in specific situations, and when an overall picture of the control properties is not needed. Accordingly, only a "local" analysis is required.

For *varying* system parameters (e.g., the excitation amplitude), on the other hand, one can be interested in shifting the homoclinic (or heteroclinic) bifurcation threshold in parameters space by the control action. Herein, the global bifurcation becomes the key point, and in principle the analysis is more complete and, consequently, more difficult. Practically this approach is aimed at modifying the regions of non homoclinic bifurcation of Fig. 4 by changing their boundaries.

By comparing the two approaches, one notes that the former is the most natural when the control is considered from a dynamical system point of view, as it involves only topological properties, while the latter is the most natural when the control is considered from an engineering point of view, because it concerns also practical aspects, such as the behaviour in parameters space, which have important consequences in terms of system response.

Actually, if the two control targets (manifolds distance and critical excitation amplitude) have the same dependence on problem parameters, the two approaches are practically equivalent, though remaining conceptually distinct. An important example is that in which we consider the homoclinic bifurcations of nearly Hamiltonian systems, detected by the Melnikov method (see previous sections), and assume the shape of the excitation as control action. We have shown in (12) and (14) that the minimum distance and the homoclinic bifurcation threshold, respectively, are both inversely proportional to the shape parameter M, so that in both cases *control entails decreasing* M.

In more general cases, such as those in which the distance is computed numerically without any specific restriction on the system, the two approaches remain distinct also from a practical point of view, and in principle they lead to practically different results, although similar qualitative behaviours are expected in common situations.

4.7 Gains and Saved Region

To quantify the improvement of control with respect to the reference (harmonic) excitation in terms of homoclinic bifurcation threshold, it is useful to introduce the so-called *gain*

$$G = \frac{\gamma_{1,\mathrm{cr}}}{\gamma_{1,\mathrm{cr}}^{h}} = \frac{1}{M}, \tag{17}$$

which is the ratio between the critical threshold with control and the critical threshold in the reference case.

With this definition, we can reformulate the control idea by saying that it consists in increasing the gain by varying γ_j and Ψ_j.

The region $\gamma^h_{1,\mathrm{cr}}(\omega) < \gamma_1 < \gamma_{1,\mathrm{cr}}(\omega) = G\gamma^h_{1,\mathrm{cr}}(\omega)$ is the zone of parameters space where arbitrary periodic excitations are theoretically effective in eliminating homoclinic chaos (with respect to the harmonic excitation), and it has been called *saved*, or controlled, region [37]. Thus, another, equivalent, point of view about the control method is that it is aimed at obtaining a saved region in the excitation amplitude-frequency parameters space.

The saved region for Helmholtz systems, whose homoclinic bifurcation threshold is depicted in Fig. 4, is reported in Fig. 5 in the case of an ω-independent value of the gain G. From this picture it is evident how the saved region is obtained by shifting the homoclinic bifurcation threshold toward higher excitation amplitudes.

4.8 Optimal Control and Optimization Problems

In the previous sections it has been established that the control method consists in varying the shape of the excitation in order (i) to increase the distance between stable and unstable manifolds, or (ii) to increase the homoclinic bifurcation threshold, or (iii) to obtain a saved region. It is natural to do these operations in a clever way, i.e., by choosing among all admissible excitations the one which *maximizes* the considered target function (d, $\gamma_{1,\mathrm{cr}}$ or the magnitude of the saved region). This point represents the second keystone of the considered control method, and is discussed in this section.

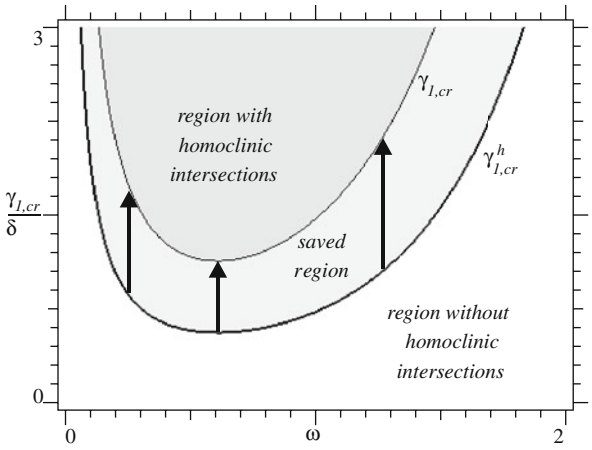


Fig. 5 The saved region for the Helmholtz oscillator [137] of Fig. 4

In both cases (i) and (ii), pursuing the objective to maximize the manifolds distance or the homoclinic critical threshold entails increasing the gain G introduced in (17). But choosing among all admissible excitations the one which maximizes the gain leads to the optimization problem

Problem P0. Maximize G by varying the Fourier coefficients γ_j and Ψ_j of the excitation,

whose solution provides the best or *optimal* excitation. Since increasing G entails decreasing $M = 1/G$, and decreasing the positive number M entails increasing

$$-M = \min_{m\in[0,2\pi]}\{h(m)\},$$

it is possible to rewrite problem *P0* in its extended form:

Problem P0'. Maximize $\min_{m\in[0,2\pi]}\{h(m)\}$ by varying the coefficients γ_j and Ψ_j of $h(m) = \sum_{j=1}^{N}\left(\frac{\gamma_j}{\gamma_1}\right)\frac{a_1(j\omega)}{a_1(\omega)}\cos(jm + \Psi_j)$.

4.8.1 Universal Optimization Problem

It is useful to reformulate problem *P0* in the following equivalent form:

Problem P1. Maximize $\min_{m\in[0,2\pi]}\{h(m)\}$ by varying the Fourier coefficients h_j and Ψ_j of $h(m) = \cos(m) + \sum_{j=2}^{N} h_j \cos(jm + \Psi_j)$.

The key advantage in passing from problem *P0* to problem *P1* is that the latter is *system-independent*, i.e. *universal*. This can be ascertain by noticing that in *P0* the system-dependent function $a_1(\omega)$ is present, while it disappears in *P1*, where the variable coefficients γ_j are "replaced" by the variable system-independent coefficients h_j.

Note that the function $a_1(\omega)$ not only is system dependent, but in general it also depends on the specific homoclinic orbit considered in systems having more than one homoclinic intersection.

Being universal, problem *P1* can be solved just once, and its solution is universally valid, namely, it is *not* necessary to solve problem *P1* for each specific system.

To illustrate the solution of problem *P1*, let us first consider the simplest case $N = 2$, i.e., the excitation is constituted by the basic harmonic plus one controlling superharmonic. This case can be handled manually [137]. In fact, the optimal choice is $\Psi_2 = 0$, and

$$\begin{aligned} M(h_2) &= -\min_{m\in[0,2\pi]}\{h(m)\} \\ &= -\min_{m\in[0,2\pi]}\{\cos(m) + h_2\cos(2m)\} \\ &= \min\left\{1 - h_2, h_2 + \frac{1}{8h_2}\right\}, \end{aligned} \tag{18}$$

so that $G(h_2) = 1/M(h_2)$. This function is depicted in Fig. 6, from which we infer that

$$G^{N=2}_{\text{optimal}} = \max_{h_2}\{G(h_2)\} = \sqrt{2}, \tag{19}$$

this maximum being attained for $h_2 = \sqrt{2}/4$.

The cases with more superharmonics are conceptually similar, but computationally more difficult, and they have been solved numerically after having proved that the condition $\Psi_j = 0$ is still optimal [37]. The results are reported in Table 1, which also shows that for $N \to \infty$ the solution $h(m)$ tends to a well defined limit constituted by a positive Dirac delta of amplitude π at $m = 0$ plus the constant function $-1/2$ (whose Fourier coefficients are in fact $h_j = 1$ and $\Psi_j = 0$). It is proved in [26, 62] that this last function actually represents the mathematical solution of the optimization problem in the case of an infinite number of superharmonics.

Table 1 The numerical results of various optimization problems with increasing number of superharmonics

N	G^N_{optimal}	h_2	h_3	h_4	h_5
2	1.4142	0.353553			
3	1.6180	0.552756	0.170789		
4	1.7321	0.673525	0.333274	0.096175	
5	1.8019	0.751654	0.462136	0.215156	0.059632
...	...	...	...	...	...
∞	2	1	1	1	1

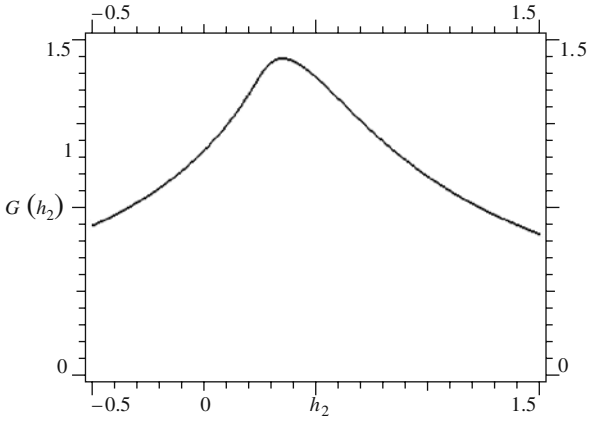


Fig. 6 The function $G(h_2)$ in the case of a single superharmonic ($N = 2$)

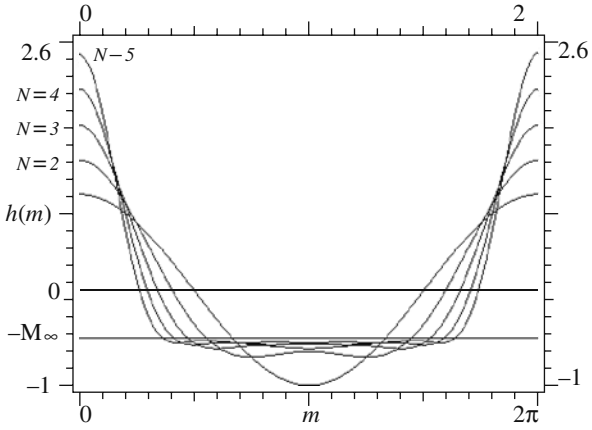


Fig. 7 Different optimal functions $h(m)$ for $N = 2, ..., 5$

Examples of optimal $h(m)$ are reported in Fig. 7 from which it is clear that the optimal solution is characterized by the presence of multiple coinciding minima. It is also possible to foresee the convergenge of that family of solutions to the limit solution for $N \to \infty$ discussed above.

4.8.2 From the Universal Optimal Solution to the Real Optimal Excitation

The link between the optimal solutions obtained in the previous point and the real optimal excitation to be applied to the structure is represented by the relations

$$\gamma_j = h_j \gamma_1 \frac{a_1(\omega)}{a_1(j\omega)}, \tag{20}$$

so that the optimal excitation can be written as

$$f_3(x,t) = \varepsilon \gamma_1 a_1(\omega) \sum_{j=1}^{N} \frac{h_j}{a_1(j\omega)} \sin(j\omega t), \tag{21}$$

which, when compared with (5) provides the optimal controlling excitation

$$\bar{f}_3(x,t) = a_1(\omega) \sum_{j=2}^{N} \frac{h_j}{a_1(j\omega)} \sin(j\omega t). \tag{22}$$

In (21) and (22) there are the "universal" coefficients h_j, which can be taken from Table 1 and do *not* need to be recalculated each time, and the function $a_1(\omega)$ which is

instead system dependent and thus changes when the control is applied to different cases.

The previous consideration further underlines one interesting, and useful, aspect of the proposed control method. In fact, upon solving the difficult part, namely the solution of the mathematical optimization problem, the operator is just left with the "technical" problem of computing the function $a_1(\omega)$ which can be "easily" done by looking at the manifolds behaviour, as discussed in the previous sub-sections. This makes the application of the control method straightforward.

Another issue arises when passing from the universal solution to practical applications. In fact, for technical reasons it may be useful to have some limitations on the optimal solution (for example, one can ask that $|\gamma_j| < |\gamma_1|$, or that $\max_{t\in[0,T]}\{\bar{f}_3(x,t)\} < \text{const.}$, etc.). These requirements can be introduced as *constraints* in the optimization problem *P1*. While not introducing special conceptual difficulties (although more sophisticated tools may be required to search for the optimal solution), these constraints are usually system dependent, so that the optimization problem turns into a *non universal* one, which is not generic and has to be solved separately for each system. This makes the application of the control method much more involved. Constrained optimization problems have been solved in [137].

4.9 Extended ("Global") and Localized ("One-Side") Application of Control

In many mechanical systems, such as symmetric oscillators, there is more than one homoclinic bifurcation which is relevant for complex dynamics and needs to be controlled (i.e., eliminated or shifted in parameter space).

In this situation, the presence of two or more homoclinic orbits permits a choice between different control strategies. Indeed, we can:

1. control *only* one (e.g., the right *or* the left in symmetric systems) homoclinic bifurcation, irrespective of what happens to the others,
2. control *simultaneously* two or more of them (e.g., the right *and* the left in symmetric systems),
3. control *simultaneously* all of them.

The choice 1 (called "one-side control" [37]) is aimed at obtaining a topologically localized control, and can be pursued by applying the points of the previous section separately to each homoclinic loop. Of course, in this case we will have an optimal excitation for each homoclinic intersection to be removed, and since these are system- and orbit-dependent, there is no hope that the optimal excitation able to remove one homoclinic intersection also succeeds in removing the other homoclinic intersections (indeed, the non-controlled homoclinic bifurcation threshold is usually lowered).

The choice 2 is aimed at controlling an extended part of the phase space, while the choice 3 is aimed at controlling the whole phase space, and thus it has been

called "global control" [37]. They require further considerations, as a consequence of the fact that in this case we want to control *simultaneously* more homoclinic intersections, i.e., we have to spread the resources available for optimization over different undesired events.

The first characteristic property of this case is that the underlying optimization problem is *no longer universal*, apart from the special case of symmetric systems with two symmetric homoclinic loops. Unfortunately, things are even more complex, and different approaches/strategies are possible. A detailed description of the whole scenario is reported in [139] with reference to the Helmholtz – Duffing oscillator. Some main results are summarized in the following.

4.9.1 Global Control of Gains

Aiming at controlling simultaneously more homoclinic bifurcations, we have to monitor the control target (distance d or homoclinic bifurcation threshold $\gamma_{1,\mathrm{cr}}$) of each homoclinic loop. From a mathematical point of view this goal can be obtained by controlling the minimum value among all bifurcations, which guarantees that all critical thresholds will not be reduced.

Since both the distance and the homoclinic bifurcation thresholds are proportional to the gain G, the first possible approach consists of maximizing the minimum gain, which yields the following optimization problem:

Problem P2. Maximize $G = \min\{G^1,G^2,...,G^L\}$ by varying the Fourier coefficients γ_j and Ψ_j of the excitation.

In the problem *P2* we have that $G^k = 1/M^k$ and

$$M^k = -\min_{m\in[0,2\pi]} \left\{ \sum_{j=1}^{N} \left(\frac{\gamma_j}{\gamma_1} \right) \frac{a_1^k(j\omega)}{a_1^k(\omega)} \cos\left(jm + \Psi_j\right) \right\}, \tag{23}$$

where $k = 1,2,..,L$ is the number of homoclinic bifurcations to be controlled simultaneously, and $a_1^k(\omega)$ is the function corresponding to the k-th homoclinic loop. Note that in general all M^ks are distinct from each other.

The (unavoidable) presence of the functions $a_1^k(\omega)$ in problem *P2* shows that it is system-dependent.

In [139], where only the case $L = 2$ is considered, this problem has been named "global control *without* symmetrization," because the homoclinic bifurcation thresholds $\gamma_{1,\mathrm{cr}}^k$, which are distinct in the reference case, remain distinct also in the optimal case. In the same paper the optimization problem has been solved in some specific cases, showing that

1. the optimal gain is much lower than that obtained in the "one-side control" (see Table 1);
2. at the optimum, the condition $G^1 = G^2$ is attained.

Roughly speaking, property 1 is a consequence of the fact that now the control "efforts" are spread over more homoclinic orbits, instead of being concentrated on a single one. This property is expected to be general, contrary to property 2 which is expected not to be satisfied, especially in the presence of more homoclinic loops.

4.9.2 Global Control of Homoclinic Bifurcation Thresholds

In the previous strategy the homoclinic bifurcation thresholds remain distinct after the application of control. It could be desirable to have a control providing, or attempting to provide, the same homoclinic bifurcation thresholds, even at the price of reducing some of them with respect to the values for harmonic excitation. This consideration leads to another optimization problem:

Problem P3. Maximize $\gamma_{1,\mathrm{cr}} = \min\left\{\gamma^1_{1,\mathrm{cr}}, \gamma^2_{1,\mathrm{cr}}, ..., \gamma^L_{1,\mathrm{cr}}\right\}$ by varying the Fourier coefficients γ_j and Ψ_j of the excitation,

where

$$\gamma^k_{1,\mathrm{cr}} = \frac{\delta a^k_0}{a^k_1(\omega)} \frac{1}{M^k}. \tag{24}$$

Note that, contrary to the previous case, now also the different expressions a^k_0 are taken into account – and this further stresses how also this case is system-dependent. In [139] this problem has been called "global control *with* symmetrization."

For increasing control "resources," the optimization problem works in the following way. It increases the lowest value of $\gamma^k_{1,\mathrm{cr}}$, say $\gamma^3_{1,\mathrm{cr}}$, until it reaches the subsequent lowest value (which could be lowered – with respect to the harmonic case – during this first phase), say $\gamma^5_{1,\mathrm{cr}}$. Then, it increases simultaneously $\gamma^3_{1,\mathrm{cr}}$ and $\gamma^5_{1,\mathrm{cr}}$, until they reach the subsequent lowest value (again, possibly decreased), and so on. This process ends when:

1. The "resources" for control finish *after* having equalized all the $\gamma^k_{1,\mathrm{cr}}$. This case has been named "global control with *achieved* symmetrization" in [139], and the optimal solution is characterized by $\gamma^1_{1,\mathrm{cr}} = \gamma^2_{1,\mathrm{cr}} = ... = \gamma^L_{1,\mathrm{cr}}$, which represents a desired circumstance.
2. The "resources" for control finish *before* having equalized all the $\gamma^k_{1,\mathrm{cr}}$. This case has been named "global control with *pursued* symmetrization" in [139].

Of course, the former case is expected when the homoclinic bifurcations of the different loops under harmonic excitation are close to each other, as occurs, for example, for weakly asymmetric systems.

The latter case, on the other hand, is expected when the reference homoclinic bifurcations are very different from each other; a limit situation is represented by the case in which the reference homoclinic bifurcation thresholds are so different that control uses all its resources to increment the lowest value, so that in this case this approach coincides with the "one-side control."

4.10 On the Application of the Control Methods to Archetypal Single-d.o.f. Systems

In the previous sub-sections, a phenomenological analysis of the homoclinic and heteroclinic bifurcations of the invariant manifolds of a saddle triggering the occurrence of complex events in dynamical systems has been presented for a generic, periodically excited, oscillator, along with the basic idea and the objective of their control, and some ensuing optimization problems.

The next issue consists in applying the control method to specific nonlinear oscillators. This has been the subject of a number of research papers [62, 63, 107, 135–146] which the reader is referred to for the details of successful application. Herein, we just aim at giving an overview of some main characterizing features of the considered systems. In this respect, it is important to refer to a suitable number of archetypal oscillators with a well-defined mathematical character and such to reliably describe the physical behaviour of real systems in *engineering science* and *mechanics*. Of course, their presentation and discussion has to be suitably organized. Indeed, nonlinear oscillators can be classified based on different viewpoints, according to whether the focus is preferably on dynamical or control aspects, while paying attention – in any case – to the characterizing features of physical world applications and to the associated computational aspects.

Herein, with the aim of stressing the mechanics environment wherein this chapter – devoted to exploiting system dynamics for control purposes – has nurtured, we choose to assume a basically *mechanical* viewpoint with the associated *dynamic* phenomena in the background, by highlighting those features of the considered systems which entail more meaningful *control* aspects. Within this overall scheme, systems are categorized according to different features, namely whether they are (i) smooth or non-smooth, whether they are characterized by (ii) single-well or multi-well potentials, by (iii) softening or hardening behaviour, by (iv) symmetric or asymmetric characteristics, also considering the occurrence or non-occurrence of a system natural frequency, and whether the technical interest is towards their (v) transient or steady response. Of course the necessary attention is also payed to the characterizing mechanical phenomena to be possibly controlled.

In terms of formulation and implementation of the dynamically based control strategy, a second viewpoint pays due attention to those *geometrical* and *computational* aspects which play a fundamental role in describing and understanding the system dynamics and in evaluating the performance of the control procedure, respectively. In this respect, the most meaningful alternatives are concerned, on one side, with dealing with either (i) homoclinic or heteroclinic manifolds, and with controlling either (ii) overall or local bifurcational events; on the other side, with the possibility to develop an (i) approximate or exact bifurcation analysis, and to implement a (ii) mainly theoretical or purely numerical control strategy. Of course, strong cross-correlations do occur amongst several of the above mentioned classification criteria. They are however believed to maintain their own interest in the aim of schematically organizing a quite rich and involved scenario of mechanical systems as well as of comparing the respective dynamic behaviours.

4.10.1 Smooth vs. Non-smooth Systems

Smoothness or non-smoothness are a basic property of mechanical systems with important consequences as regards their mathematical description and dynamical characterization. Thus, they are assumed as a first criterion to distinguish between two broad classes of systems.

Due to the occurrence of discontinuities in either the displacement or the velocity, non-smooth systems exhibit a further enriched pattern of response classes and local/global bifurcations with respect to the already involved scenario of smooth systems, along with some main novelties in the dynamics features.

In terms of the proposed control, the main difference is that in smooth systems the homoclinic bifurcation cannot be computed analytically, not even for the homoclinic orbits of hilltop saddles. The only analytical results are obtained by perturbative approaches, as the Melnikov method [148] or the method illustrated in Sect. 4.3. For the piece-wise linear systems, instead, the homoclinic bifurcation threshold can be computed exactly, and this allows a more detailed application of control.

A number of fully nonlinear oscillators with continuous-type nonlinearities, of either exact (such as the mathematical pendulum) or approximate, polynomial-kind (such as the Duffing [136, 138], Helmholtz [137], Helmholtz-Duffing [139] oscillators, a micromechanical system [142]) nature, have been considered. The latter mostly – though not exclusively – describe the nonlinear single-mode, non-internally resonant, dynamics of a number of smooth infinite-dimensional systems of interest in structural and engineering mechanics.

Two piecewise linear oscillators have been dealt with, i.e. the impacting inverted pendulum [26, 62, 135, 107] and the rocking rigid block [63, 140, 141], to be intended, respectively, as paradigmatic models for non-smooth nonlinear dynamics of mathematical/mechanical significance or more practical interest.

4.10.2 Single-Well vs. Multi-Well Potentials

Depending on the energy imparted to the system through the initial conditions or the excitation amplitude, as well as on the values of other control parameters, two main issues are of considerable practical importance for a mechanical system, namely (i) the possibility to have the system actually working within the range of operating conditions which it has been designed for, and (ii) the loss of technical performance or structural integrity ensuing from the system getting off its foreseen operational regime.

In dynamical terms, it is possible to restrict ourselves to systems characterized by the existence of either a single-well or a (possibly) multi-well potential, for which two different critical situations may occur, namely,

1. the motion does develop on a *small scale* in the phase space, being actually restricted within a sole, i.e. the foreseen, potential well, however with dynamical features other than the wanted ones, as it may occur due to any in-well bifurcation of the system dynamics to a different periodic regime or to an aperiodic regime;

2. the motion does develop on a *large scale* in the phase space, i.e. also beyond the invariant manifolds of the hilltop saddle which delimit the working potential well and organize the whole system dynamics.

The occurrence of either a single-well or a multiple well has a considerable influence as regards various dynamics, bifurcation, and control issues.

4.10.3 Softening vs. Hardening Systems

The distinction between hardening and softening systems reflects the intrinsic mathematical/mechanical character of the system constitutive law and the associated dynamic effects occurring for finite amplitude motions. It mostly pertains to the large scale motions arising upon the global bifurcation of the hilltop saddle manifolds, which can occur in both single-well and multiple-well systems, and it is actually concerned with the kind of out-of-well attractor onto which the system settles down upon overcoming the bounding invariant manifolds, i.e. after escaping from the potential well. Again, two basically different situations can occur.

1. Though being no more restricted within the reference well, the overall system dynamics still remains *bounded*. This is the case where, after escape, the motion develops entirely within a neighbouring, bounded, well, or it wanders around, e.g., two wells, the reference and a neighbouring one. Scenarios of this kind may occur in multi-well hardening systems and are usually unpleasant from the application viewpoint, yet not destructive. Aperiodic motion scattered between adjacent wells, or the so called cross-well chaos, are indeed classes of motion pertaining to such after-escape non-destructive regimes, and are quite common in several (e.g., two-well Duffing-type [136, 138, 139]) systems. Of course, such kinds of unwanted dynamics may need to be avoided or somehow controlled. However, technical situations can be devised wherein realizing alternating dynamic regimes between the reference well and an adjacent one can even be desirable: let us think, e.g., of the shift between oscillating (in-well) and rotating (out-of-well) regimes in a planar pendulum subjected to vertical harmonic (i.e. parametric) excitation of its support [150].
2. The system dynamics escapes to "infinity". This is the case where, after escape, the motion is theoretically *unbounded*, and practically corresponds to the system settling down onto an attractor far away and completely different from the designed one or, in other terms, to the definitive system failure. Mechanical situations of reference in this respect are, e.g., the capsizing of a rolling ship in ocean waves [7, 137, 151], the overturning of a rocking rigid block under periodic acceleration of its basement [140, 141], or the pull-in phenomenon occurring in a microelectromechanical system [142] to be used as a resonator or sensor device, which corresponds to the microbeam collapse onto the electrically charged substrate. However, there are again technical situations wherein realizing a regime nominally "at infinity" is highly desirable: think, e.g., of the pull-in dynamics of the microbeam in a MEMS to be used as a switch device.

It is also worth stressing two more meaningful points. (a) Both the two different escape situations just illustrated are actually characterized, from a topological viewpoint, by a progressive erosion process of either a single in-well basin of attraction or of the whole collection of in-well basins. Yet, the degree of loss of dynamical integrity [146, 147] and the criticality of the final outcomes are considerably different in the two cases. (b) The scenario of bounded/unbounded dynamical regimes of a given system in the excitation parameters space can be influenced by the occurrence or non-occurrence of a system natural frequency, the former case naturally entailing a possibility of resonant excitation condition. This may meaningfully affect the (major or minor) system sensitivity, in terms of loss of dynamical integrity, to the variation of a leading bifurcation parameter.

4.10.4 Homoclinic vs. Heteroclinic Bifurcations

As already discussed, the occurrence of a generic underlying bifurcational event, i.e. of the global, homoclinic or heteroclinic, bifurcation of a (possibly) hilltop saddle governing the whole system dynamics (or most of it), is the necessary prerequisite for developing a generic approach to control – or design – of mechanical systems. Yet, several distinct topological situations do occur for different oscillators. By way of example, in the phase portrait one can have (i) one homoclinic orbit surrounding the unique potential well, like in Helmholtz-type softening oscillators [137], (ii) two homoclinic orbits surrounding two separate wells, joined through the intermediate hilltop saddle, like in Duffing-type hardening oscillators [136, 138, 139] – in both cases irrespective of the nature of the system being smooth or non-smooth –, or (iii) one couple of heteroclinic orbits, each one connecting the two organizing hilltop saddles, which surround the unique potential well, like in the Duffing-type softening oscillators or in the rocking block [63, 140, 141]. Of course, more involved phase portraits, characterized by the occurrence of both homoclinic and heteroclinic orbits, may also occur.

While in the case of a single (homoclinic) unperturbed manifold there is only one possible global bifurcation corresponding to the tangency of the split perturbed manifolds, in the case of two distinct (either homoclinic or heteroclinic) unperturbed manifolds the global bifurcations can occur either simultaneously or separately, depending on the values of some varying perturbation parameter. And this has very important consequences in terms of control, as we have seen above.

It is worth mentioning how, in a given system, the homo/heteroclinic bifurcation of the invariant manifolds of an hilltop saddle – i.e, of a globally organizing saddle – does represent the main bifurcational event governing the transition of system dynamics from a small scale to a large scale regime. Yet, as a rule, other local saddles (up to possibly infinitely many) also exist within each potential well, as a consequence of either the instability of a single in-well periodic solution or of the onset of a chaotic saddle born at a previous homo/heteroclinic bifurcation of the coexisting hilltop. Though being *secondary* bifurcational events, the intersections of the invariant manifolds of a local saddle may also entail meaningful changes of system dynamics, either on the small scale (i.e., in-well) regime or on the large

scale (i.e., cross-well) regime. The first one may correspond, e.g., to the transition from a locally periodic to a locally complex (in-well chaos) response. The second, usually as a consequence of the involvement in the global bifurcational event of also the hilltop saddle due to some heteroclinic connection, may correspond, e.g., to the transition from in-well to cross-well chaos which occurs according to any "turning around" topological mechanism (see, e.g., [152]).

Before addressing the control issue in various possible situations, two more *dynamical* issues, which are also representative of major mechanical differences, are mentioned.

4.10.5 Symmetric vs. Asymmetric Systems

The symmetry or asymmetry in the force-displacement relationship of the underlying mechanical system do reflect themselves in distinguishing features in the potential and the unperturbed phase space of the associated nonlinear oscillators.

This distinction deserves attention mostly from a dynamical systems point of view. Focusing on the more "dangerous" softening oscillators, let us think of the difference – under symmetric excitation – between a single-well potential with one escape direction (the smooth Helmholtz oscillator or the non-smooth monolaterally unconstrained rocking block) or with two asymmetric escape directions (the Helmholtz-Duffing oscillator), on one side, and a single-well potential with two symmetric escape directions (the Duffing oscillator or the bilaterally unconstrained rocking block), on the other side. The solely or prevailing sensitivity to escape ensuing from the asymmetry feature, respectively associated with the occurrence of either one single hilltop saddle or of two hilltop saddles of different energy level, does correspond, in the various cases, to weaker, asymmetric, mechanical features in the background.

In contrast, the symmetry entails, e.g., in softening systems – such as the Duffing or the rigid block oscillators – that the potential well is surrounded by a couple of heteroclinic orbits which become a unique heteroclinic loop connecting two equal energy hilltop saddles in the presence of imperfections, whereas in hardening systems – such as the two-well Duffing oscillator – the two homoclinic bifurcations occur simultaneously for all symmetric excitations.

Moreover, still thinking in dynamical terms, it must be underlined how, apart from some pathological situations, symmetric oscillators are structurally unstable and can be considered as particular cases of asymmetric oscillators, though they are not always a limit, in an appropriate sense, of the latter (see, e.g., the softening Helmholtz-Duffing equation in [137]).

4.10.6 Transient vs. Steady Dynamics

The dynamics, as well as the control, perspectives are quite different depending on whether the mechanical interest in the background technical application is towards the transient or the steady system response. In the problem of ship capsize, for example, it has been emphasised that short-term transient behaviour is particularly

important [153]. As a matter of fact, there are situations where a temporary escape from the potential well may be unessential for the system operating effectiveness, whereas it has to be strictly avoided in other practical situations. The transient regime is very important and governs the whole system performance in the non-linear dynamics due to short-term excitations (such as those associated with impact forces or seismic loads), while the steady state dynamics are apparently of major interest in the presence of stationary, long-term, excitations. Moreover, the transient regime has obviously as minor importance as shorter it is (like it occurs, e.g., in highly damped systems) while it becomes crucial when it is long (e.g., in highly deformable systems).

All of the previous *mechanical*, *bifurcational*, and *dynamical* distinguishing features have meaningful influences on the realizable control scenarios.

4.10.7 Overall vs. Localized Control

Among various possible categorization perspectives, the broad one in this subsection title basically distinguishes between overall and localized control, and seems to be a rather comprehensive, as well as comparatively useful, one. As a matter of fact, it strictly depends on the two previously addressed, distinct, aspects of the whole problem, namely (i) the *main* or *secondary* character of the bifurcational event of interest, and (ii) the *large* or *small* scale nature of the ensuing dynamics.

As already said, the global bifurcations of the manifolds of a hilltop saddle are main events organizing the system dynamics on a large scale. Nonetheless, in terms of control, a basic distinction does apply as to the possibility (or the interest) to control the global bifurcations, and the ensuing system dynamics, over the whole phase space or just one meaningful part of it. Yet, within this general framework, different situations must be considered.

1. The pursued distinction makes actually full sense only in systems exhibiting (at least) two separate homoclinic orbits, organized by one intermediate hilltop saddle and surrounding two clearly identified topological regions in the phase space, which correspond to the two potential wells: this is the case, e.g., of the Duffing-type (smooth or non-smooth) hardening oscillators. In such systems, one can aim at controlling only one of the two homoclinic bifurcations, which also entails controlling the dynamics in one clearly identified part of the phase space: for apparent reasons, this topologically localized control strategy has been termed "one-side control" in Sect. 4.9. Alternatively, one can aim at controlling both the two homoclinic bifurcations, which clearly corresponds to controlling, on average, the dynamics in the whole phase space: this topologically overall control strategy naturally has been termed "global control" in Sect. 4.9.
2. A similar distinction applies to the escape systems which exhibit two separate heteroclinic orbits connecting two organizing hilltop saddles and surrounding a unique potential well: this is the case of the Duffing-type softening oscillator or of the rocking block. Yet, herein, the distinction is considerably weaker from the application viewpoint. Indeed, also in such systems one can aim at

controlling either only one of the two heteroclinic bifurcations, thus realizing a "one-side control," or both heteroclinic bifurcations, thus realizing a "global control." However, the former strategy no more entails controlling the dynamics in one clearly identified part of the phase space or, in other terms, it no more produces a topologically localized control. What actually happens is that the perturbed heteroclinic loop responsible for complex dynamic phenomena (such as a chaotic saddle and the ensuing horseshoe dynamics) actually breaks down, but the heteroclinic intersection of the non-controlled orbit still allows for the penetration of tongues of the basin of attraction of the infinity into the in-well bounded basins, which may ultimately result in a much earlier system failure.

3. The third case of interest, within the large scale dynamics ensuing from one main bifurcational event, is concerned with all escaping situations characterized by only one homoclinic orbit, which delimits a clearly defined part of the phase space: this is the case of all Helmholtz-type softening oscillators [137]. For consistency with the previously adopted terminology, controlling the solely possible homoclinic bifurcation corresponds to realizing a "one-side control," which in this case actually entails controlling the sole topologically meaningful part of the system phase space.

The second distinction between overall and localized control is concerned with the secondary bifurcational events possibly associated with local saddles, and with the ensuing system dynamics. It refers to the possibility, and the interest, to control not so much the global bifurcations of the invariant manifolds of the hilltop saddle organizing the whole dynamics as those of the manifolds of some non-hilltop saddle governing the dynamics within a restricted area of the phase space and making sense in specific technical situations. Yet, at least two schematically different situations may occur, again.

1. If the secondary event entails the onset of a localized – though possibly complex – dynamics on a *small scale* (e.g., in-well chaos), only the "one-side control" (from the large scale dynamics perspective mentioned at last point 3) can of course be applied.
2. If, on the contrary, due to the involvement of also the hilltop saddle through an heteroclinic connection, the secondary bifurcational event entails a *large scale* dynamics (think, e.g., of the transition from a single-well to a cross-well chaos in a specific hardening Duffing oscillator), one can again distinguish between different control strategies (i.e., "global" vs. "one-side") along the lines previously discussed.

Yet, there is a fundamental difference with respect to the hilltop saddle manifolds bifurcation. Indeed, while in this case there can be a very useful chance to develop an *analytical*, either exact or approximate, treatment [63], no analytical tools are usually available to detect specific homo/heteroclinic bifurcations of other saddles playing an only locally dominant role in the dynamics. In such cases, the lack of mathematical tools requires adopting a fully *numerical* approach, wherein both the invariant manifolds and their bifurcations are detected through computer simula-

tions [136], with obvious restrictions in terms of both generality and applicability of the procedure.

4.10.8 System-Independent vs. System-Dependent Controls

With reference to the control scenario on the large scale, it is also worth mentioning some further, very important, control aspects associated with realizing a "one-side" or "global" control as well as with considering hardening symmetric or asymmetric systems.

They basically consist in the occurrence of meaningful generic (or sub-universal) features of the mathematical problems of optimization governing the cases of "one-side control" of whatever (softening or hardening) systems and of "global control" of hardening symmetric systems, which also entail very desirable system-independent solutions. This is in contrast with the non-generic optimal "global" control of hardening asymmetric systems, whose solutions are indeed system-dependent and may actually aim at different outcomes of the controlled dynamics, which are strictly associated with the existing asymmetry features.

4.10.9 Finite- vs. Infinite-Dimensional Systems

Finally, the very important issue of the finite- vs. infinite- dimensionality of the system has to be mentioned. The matter was already addressed in Sect. 1, when dwelling on the reliability of the reduced order models usually employed for analyzing the nonlinear dynamics and control of mechanical and structural systems.

In the framerwork of the present control method, from a technical point of view, detecting the unwanted global bifurcation may be quite difficult in infinite-dimensional systems, both analytically and numerically, especially because the manifolds are no longer 1D.

Referring to a buckled beam with variable boundary conditions, it has been shown [143] how the (minimal) reduced order model is able to provide the same homoclinic bifurcation threshold and the same optimal control as the actual infinite-dimensional system in the background, although this is accomplished in an exact way or with a practically acceptable approximation depending on some problem characteristics. However, in all cases, the correct infinite-dimensional analysis may add useful new resources for control purposes, thus enlarging the potentiality of the method. For example, in infinite-dimensional systems one can determine and use also the "optimal" *spatial* shape of the controlling excitation, in addition to the optimal *temporal* one solely obtainable in single d.o.f. systems [143].

References

1. J. M. T. Thompson and R. Ghaffari. Chaos after period-doubling bifurcations in the resonance of an impact oscillator. *Phys. Lett*., **91A**, 5 (1982).
2. C. Grebogi, E. Ott, and J. A. Yorke. Chaotic attractors in crisis. *Phys. Rev. Lett*., **48**(22), 1507 (1982).

3. J. M. T. Thompson. Complex dynamics of compliant off-shore structures. *Proc. R. Soc. Lond. A*, **387**, 407 (1983).
4. J. M. T. Thompson and H. B. Stewart *Nonlinear Dynamics and Chaos*. Wiley, Chichester (1986 – Second Edition 2002).
5. C. Grebogi, E. Ott, and J. A. Yorke. Crises, sudden changes in chaotic attractors, and transient chaos. *Phys. D*, **7**(1–3), 181 (1983).
6. C. Grebogi, E. Ott, and J. A. Yorke. Are three-frequency quasiperiodic orbits to be expected in typical nonlinear dynamical systems? *Phys. Rev. Lett*., **51**(5), 339 (1983).
7. J. M. T. Thompson. Chaotic phenomena triggering the escape from a potential well. *Proc. R. Soc. Lond. A*, **421**, 195 (1989).
8. S. W. McDonald, C. Grebogi, E. Ott, and J. A. Yorke. Fractal basin boundaries. *Phys. D*, **17**(2), 125 (1985).
9. E. Ott, C. Grebogi, and J. A. Yorke. Controlling chaos. *Phys. Rev. Lett*., **64**(11), 1196 (1990).
10. G. R. Chen and X. Dong. *From Chaos to Order: Methodologies, Perspectives and Applications*. World Scientific, Singapore (1998).
11. G. R. Chen. *Controlling Chaos and Bifurcations in Engineering Systems*. CRC Press, Boca Raton, FL (1999).
12. K. Judd, A. I. Mees, K. L. Teo, and T. Vincent. *Control of Chaos*. Birkhauser, Boston, MA (1997).
13. A. L. Fradkov and A. Yu Pogromsky. *Introduction to Control of Oscillations and Chaos*. World Scientific, Singapore (1998).
14. T. Kapitaniak. *Controlling Chaos*. Academic Press, London (1996).
15. M. Lakshmanan and K. Murali. *Chaos in Nonlinear Oscillators: Controlling and Synchronization*. World Scientific, Singapore (1996).
16. H.-G. Schuster (Ed.). *Handbook of Chaos Control: Foundations and Applications*. Wiley VCH, Weinheim (1999).
17. R. Chacon. *Control of Homoclinic Chaos by Weak Periodic Perturbations*, Series on Nonlinear Science, Series A. World Scientific, Singapore, vol. 55 (2005).
18. T. Arecchi, S. Boccaletti, M. Ciofini, C. Grebogi, and R. Meucci (Eds.). *Int. J. Bifurcat. Chaos*, **8**, 1739–1742 (1998).
19. G. R. Chen and M. J. Ogorzalek (Eds.). *Int. J. Bifurcat. Chaos*, **10** (2000).
20. G. R. Chen (Ed.). *Int. J. Bifurcat. Chaos*, **12** (2002).
21. W. L. Ditto and K. Showalter (Eds.). *Chaos*, **7**, 509–687 (1997).
22. M. Hasler and J. Vandevalle (Eds.). *Int. J. Circ. Theor. Appl*., **27**, 525–526 (1999).
23. T. Kapitaniak (Ed.). *Chaos Solitons Fractals*, **8** (1997).
24. M. Kennedy and M. Ogorzalek (Eds.). *IEEE Trans. Circ. Syst*., **44** (1997).
25. H. Nijmeijer (Ed.). *Syst. Control Lett*. **31**, 259–262 (1997).
26. S. Lenci and G. Rega (Eds.). *Phil. Trans. R. Soc. A*, **364** (2006).
27. T. Arecchi, S. Boccaletti, M. Ciofini, R. Meucci, and C. Grebogi. The control of chaos: theoretical schemes and experimental realizations. *Int. J. Bifurcat. Chaos*, **8**(8), 1643 (1998).
28. B. Blazejczyk, T. Kapitaniak, J. Wojewoda, and J. Brindley. Controlling chaos in mechanical systems. *Appl. Mech. Rev*., **46**(7), 385 (1993).
29. S. Boccaletti, C. Grebogi, Y. C. Lai, H. Mancini, and D. Maza. The control of chaos: theory and applications. *Phys. Rep*., **329**, 103 (2000).
30. G. R. Chen and X. Dong. From chaos to order – Perspectives and methodologies in controlling chaotic nonlinear dynamical systems, *Int. J. Bifurcat. Chaos*, **3**(6), 1363 (1993).
31. G. Chen. Introduction to chaos control and anti-control. In T. P. Leung and H.-S. Qin, editors, *Advanced Topics in Nonlinear Control Systems*. World Scientific, Singapore, pp. 193–245 (2001).
32. M. Z. Ding, E. J. Ding, W. L. Ditto, B. Gluckman, V. In, J. H. Peng, M. L. Spano, and W. M. Yang. Control and synchronization of chaos in high dimensional systems: review of some recent results, *Chaos*, **7**(4), 644 (1997).
33. A. L. Fradkov. Chaos control bibliography (1997–2000). Russian Systems and Control Archive (RUSYCON), www.rusycon.ru/chaoscontrol. html (2000).

34. A. L. Fradkov and R. L. Evans. Control of chaos: methods and applications in engineering. *Ann. Rev. Control*, **29**(1), 33 (2005).
35. A. L. Fradkov, R. L. Evans, and B. R. Andrievski. Control of chaos: methods and applications in mechanics. *Phil. Trans. R. Soc. A*, **364**, 2279 (2006).
36. J. F. Lindner and W. L. Ditto. Removal, suppression, and control of chaos by nonlinear design, *Appl. Mech. Rev*., **48**(12), 795 (1995).
37. G. Rega and S. Lenci. Bifurcations and chaos in single-d.o.f. mechanical systems: exploiting nonlinear dynamics properties for their control. In A. Luongo, editor, *Recent Research Developments in Structural Dynamics*. Transworld Research Network, Kerala, pp. 331–369 (2003).
38. G. Rega and S. Lenci. Introduction**.** *Phil. Trans. Royal Society A*, **364**, 2269 (2006).
39. D. J. Gauthier. Resource letter: CC-1: controlling chaos. *Am. J. Phys.*, **71**(8), 750 (2003).
40. S. W. Shaw. The suppression of chaos in periodically forced oscillators. In W. Schiehlen, editor, *Nonlinear Dynamics in Engineering Systems*. Springer-Verlag, Berlin, pp. 289–296 (1990).
41. R. Lima and M. Pettini. Suppression of chaos by resonant parametric perturbations. *Phys. Rev. A*, **41**, 726 (1990).
42. F. Casas and C. Grebogi. Control of chaotic impacts. *Int. J. Bifurcat. Chaos*, **7**(4), 951 (1997).
43. M. Ding, E. Ott, and C. Grebogi. Controlling chaos in a temporally irregular environment. *Phys. D*, **74**, 386 (1994).
44. E. N. Macau and C. Grebogi. Control of chaos and its relevance to spacecraft steering. *Phil. Trans. R. Soc. A*, **364**, 2462 (2006).
45. T. Shinbrot, W. Ditto, C. Grebogi, E. Ott, M. Spano, and J. A. Yorke. Using the sensitive dependence of chaos (the "butterfly effect") to direct trajectories in an experimental chaotic system. *Phys. Rev. Lett*., **68**(19), 2863 (1992).
46. T. Shinbrot, E. Ott, C. Grebogi, and J. A. Yorke. Using chaos to direct trajectories to targets. *Phys. Rev. Lett*., **65**(26), 3215 (1990).
47. T. Shinbrot, E. Ott, C. Grebogi, and J. A. Yorke. Using small perturbations to control chaos. *Nature*, **363**, 411 (1993).
48. K. Yagasaki and T. Uozumi. A new approach for controlling chaotic dynamical systems. *Phys. Lett. A*, **238**, 349 (1998).
49. K. Yagasaki and T. Uozumi. Controlling chaos using nonlinear approximations and delay coordinate embedding. *Phys. Lett. A*, **247**, 129 (1998).
50. B. Epureanu, S. T. Trickey, and E. Dowell. Controlling chaos using nonlinear approximations and delay coordinate embedding. *Nonlinear Dyn*., **15**, 191 (1998).
51. B. Epureanu and E. Dowell. On the optimality of the Ott–Grebogi–Yorke control scheme. *Phys. D*, **116**, 1 (1998).
52. D. L. Hill. On the control of chaotic dynamical systems using nonlinear approximations. *Int. J. Bifurcat. Chaos*, **11**(1), 207 (2001).
53. D. L. Hill. On the control of high dimensional chaotic dynamical systems using nonlinear approximations. *Int. J. Bifurcat. Chaos*, **11**(6), 1753 (2001).
54. M. K. Sifakis and S. J. Elliott. Strategies for the control of chaos in a Duffing–Holmes oscillator. *Mech. Syst. Signal Process*., **14**, 987 (2000).
55. J. Singer, Y. Z. Wang, and H. H. Bau. Controlling a chaotic system. *Phys. Rev. Lett*., **66**(9), 1123 (1991).
56. K. Pyragas. Continuous control of chaos by self-controlling feedback. *Phys. Lett. A*, **170**, 421 (1992).
57. G. Filatrella, G. Rotoli, and M. Salerno. Suppression of chaos in the perturbed Sine–Gordon system by weak periodic signals. *Phys. Lett. A*, **178**, 81 (1993).
58. J. Guémez, J. M. Gutiérrez, A. Iglesias, and M. A. Matias. Suppression of chaos trough changes in the system variables: transient chaos and crises. *Phys. D*, **79**, 164 (1994).
59. S. Lenci. On the suppression of chaos by means of bounded excitations in an inverted pendulum. *SIAM J. Appl. Math*., **58**(4), 1116 (1998).
60. A. Isidori. *Nonlinear Control Systems*. Springer-Verlag, Rome (1995).

61. O. C. Pinto and P. Gonçalves. Non-linear control of buckled beams under step loading. *Mech. Syst. Signal Process*., **14**, 967 (2000).
62. S. Lenci and G. Rega. A procedure for reducing the chaotic response region in an impact mechanical system. *Nonlinear Dyn*., **15**(4), 391 (1998).
63. S. Lenci and G. Rega. A unified control framework of the nonregular dynamics of mechanical oscillators. *J. Sound Vibr*., **278**, 1051 (2004).
64. S. Boccaletti, A. Farini, E. J. Kostelich, and T. Arecchi. Adaptive targeting of chaos. *Phys. Rev. E*, **55**(5), 4845 (1997).
65. C. M. Bird and P. J. Aston. Targeting in the presence of noise. *Chaos Solitons Fractals*, **9**(1/2), 251 (1998).
66. T. Shinbrot. Chaos: unpredictable yet controllable? *Nonlinear Sci. Today*, **38**(2), 1 (1993).
67. E. Ott and M. Spano. Controlling chaos. *Phys. Today*, **48**, 34 (1995).
68. E. Ott. *Scholarpedia*, **1**(8), 1699 (2006).
69. F. Takens, Detecting strange attractors in turbulence. In D. A. Rand and L. S. Young, editors, *Proceedings of the Symposium on Dynamical Systems and Turbulence*. Springer, Berlin (1981).
70. P. So, E. Ott, T. Sauer, B. J. Gluckman, C. Grebogi, and S. J. Schiff. Extracting unstable periodic orbits from chaotic time series data. *Phys. Rev. E*, **55**, 5398 (1997).
71. D. Pierson and F. Moss. Detecting periodic unstable points in noisy chaotic and limit cycle attractors with applications to biology. *Phys. Rev. Lett*., **75**, 2124 (1995).
72. B. R. Hunt and E. Ott. Optimal periodic orbits of chaotic systems. *Phys. Rev. Lett*., **76**, 2254 (1996).
73. F. J. Romeiras, E. Ott, C. Grebogi, and W. P. Dayawansa. Controlling chaotic dynamical systems. *Phys. D*, **58**, 165 (1992).
74. U. Dressler and G. Nitsche. Controlling chaos using time delay coordinates. *Phys. Rev. Lett*., **68**(1), 1 (1992).
75. P. So and E. Ott. Controlling chaos using time delay coordinates via stabilization of periodic orbits. *Phys. Rev. E*, **51**(4), 2955 (1995).
76. J. E. S. Socolar, D. W. Sukow, and D. J. Gauthier. Stabilizing unstable periodic orbits in fast dynamical systems. *Phys. Rev. E*, **50**, 3245 (1994).
77. E. Kostelich, C. Grebogi, E. Ott, and J. A. Yorke. Higher-dimensional targeting. *Phys. Rev. E*, **47**, 305 (1993).
78. E. M. Bollt and J. D. Meiss. Targeting chaotic orbits to the Moon through recurrence. *Phys. Lett. A*, **204**, 373 (1995).
79. C. Schroer and E. Ott. Targeting in Hamiltonian systems that have mixed regular/chaotic phase spaces. *Chaos*, **7**, 512 (1997).
80. C. J. Begley and L. N. Virgin. On the OGY control of an impact-friction oscillator. *J. Vibr. Control*, **7**, 923 (2001).
81. N. Packard, J. Crutchfield, D. Farmer, and R. Shaw. Geometry from a time series. *Phys. Rev. Lett*., **45**, 712 (1980).
82. E. Ott, T. Sauer, and J. A. Yorke. *Coping with Chaos: Analysis of Chaotic Data and the Explanation of Chaotic Systems*. Wiley Interscience, New York (1994).
83. K. Yagasaki and Y. Tochio. Experimental control of chaos by modifications of delayed feedback. *Int. J. Bifurcat. Chaos*, **11**, 3125 (2001).
84. P. V. Bayly and L. N. Virgin. Practical considerations in the control of chaos. *Phys. Rev. E*, **50**, 604 (1994).
85. B. Hübinger, R. Doerner, W. Martienssen, M. Herdering, R. Pitka, and U. Dressler. Controlling chaos experimentally in systems exhibiting large effective Lyapunov exponents. *Phys. Rev. E*, **50**, 932 (1994).
86. T. Ritz, A. S. Z. Schweinsberg, U. Dressler, R. Doerner, B. Hübinger, and W. Martienssen. Chaos control with adjustable control times. *Chaos Solitons Fractals*, **8**(9), 1559 (1997).
87. R. J. DeKorte, J. C. Schouten, and C. M. van den Bleek. Experimental control of a chaotic pendulum with unknown dynamics using delay coordinates. *Phys. Rev. E*, **52**(4), 3358 (1995).

88. J. Starrett and R. Tagg. Control of a chaotic parametrically driven pendulum. *Phys. Rev. Lett*., **74**(11), 1974 (1995).
89. K. Yagasaki and T. Uozumi. Controlling chaos in a pendulum subjected to feedforward and feedback control. *Int. J. Bifurcat. Chaos*, **7**, 2827 (1997).
90. D. J. Christini, J. J. Collins, and P. S. Linsay. Experimental control of high-dimensional chaos: the driven double pendulum. *Phys. Rev. E*, **54**(5), 4824 (1996).
91. G. L. Baker. Control of the chaotic driven pendulum. *Am. J. Phys*., **63**, 832 (1995).
92. D. L. Xu and S. R. Bishop. Steering dynamical trajectories to target a desired state. *Chaos Solitons Fractals*, **4**, 1931 (1994).
93. S. R. Bishop, D. L. Xu, and M. J. Clifford. Flexible control of the parametrically excited pendulum. *Proc. R. Soc. Lond. A*, **452**, 1789 (1996).
94. K. Yagasaki and S. Yamashita. Controlling chaos using nonlinear approximations for a pendulum with feedforward and feedback control. *Int. J. Bifurcat. Chaos*, **9**(1), 233 (1999).
95. R. Wang and Z. Jing. Chaos control of chaotic pendulum system. *Chaos Solitons Fractals*, **21**, 201 (2004).
96. A. Alasty and H. Salarieh. Nonlinear feedback control of chaotic pendulum in presence of saturation effect. *Chaos Solitons Fractals*, **31**, 292 (2007).
97. F. H. I. Pereira-Pinto, A. M. Ferreira, and M. A. Savi. Chaos control in a nonlinear pendulum using a semi-continuous method. *Chaos Solitons Fractals*, **22**, 653 (2004).
98. F. H. I. Pereira-Pinto, A. M. Ferreira, and M. A. Savi. State space reconstruction using extended state observers to control chaos in a nonlinear pendulum. *Int. J. Bifurcat. Chaos*, **15**(12), 4051 (2005).
99. W. L. Ditto, S. N. Rauseo, and M. L. Spano. Experimental control of chaos. *Phys. Rev. Lett*., **65**(26), 3211 (1990).
100. F. C. Moon. *Chaotic and Fractal Dynamics*. Wiley, New York (1992).
101. F. C. Moon, M. A. Johnson, and W. T. Holmes. Controlling chaos in a two-well oscillator. *Int. J. Bifurcat. Chaos*, **6**(2), 337 (1996).
102. E. R. Hunt. Stabilizing high-periodic orbits in a chaotic system: the diode resonator. *Phys. Rev. Lett*., **67**(15), 1953 (1991).
103. U. Dressler, T. Ritz, A. Schenck zu Schweinsberg, R. Doerner, B. Hübinger, and W. Martienssen. Tracking unstable periodic orbits in a bronze ribbon experiment. *Phys. Rev. E*, **51**, 1845 (1995).
104. M. Ding, W. Yang, V. In, W. L. Ditto, M. L. Spano, and B. Gluckman. Controlling chaos in high dimensions: theory and experiment. *Phys. Rev. E*, **53**, 4334 (1996).
105. M. Ding, E. Ott, and C. Grebogi. Crisis control: preventing chaos-induced capsizing of a ship. *Phys. Rev. E*, **50**, 4228 (1994).
106. E. R. Hunt. Keeping chaos at bay. *IEEE Spectr*., **30**, 32 (1993).
107. S. Lenci and G. Rega. Numerical control of impact dynamics of an inverted pendulum through optimal feedback strategies. *J. Sound Vibr*., **236**(3), 505 (2000).
108. J. Kalagnanam. Controlling chaos: the example of an impact oscillator. *J. Dyn. Syst. Meas. Control*, **116**, 557 (1994).
109. A. B. Nordmark. Non-periodic motion caused by grazing incidence in an impact oscillator. *J. Sound Vibr*., **145**, 279 (1991).
110. S. L. T. de Souza and I. L. Caldas. Controlling chaotic orbits in mechanical systems with impacts. *Chaos Solitons Fractals*, **19**, 171 (2004).
111. S. R. Bishop, D. J. Wagg, and D. Xu. Use of control to maintain period-1 motions during wind-up or wind-down operations of an impacting driven beam. *Chaos Solitons Fractals*, **9**(1/2), 261 (1998).
112. U. Galvanetto. Flexible control of chaotic stick-slip mechanical systems. *Comput. Meth. Appl. Mech. Eng*., **190**(46–47), 6075 (2001).
113. E. Gutiérrez and D. K. Arrowsmith. Control of a double impacting mechanical oscillator using displacement feedback. *Int. J. Bifurcat. Chaos*, **14**(9), 3095 (2004).
114. S. Chatterjee, A. K. Mallik, and A. Ghosh. On impact dampers for non-linear vibration systems. *J. Sound Vibr*., **187**, 403 (1995).

115. S. Chatterjee, A. K. Mallik, and A. Ghosh. Impact dampers for controlling self-excited oscillation. *J. Sound Vibr*., **193**, 1003 (1995).
116. S. L. T. de Souza, I. L. Caldas, R. L. Viana, J. M. Balthazar, and R. M. L. R. F. Brasil. Impact dampers for controlling chaos in systems with limited power supply. *J. Sound Vibr*., **279**, 955 (2005).
117. F. C. Moon, A. J. Reddy, and W. T. Holmes. Experiments in control and anti-control of chaos in a dry friction oscillator. *J. Vibr. Control*, **9**, 387 (2003).
118. U. Feudel, C. Grebogi, L. Poon, and J. A. Yorke. Dynamical properties of a simple mechanical system with a large number of coexisting periodic attractors. *Chaos Solitons Fractals*, **9**(1–2), 171 (1998).
119. H. N. Agiza. Controlling chaos for the dynamical system of coupled dynamos. *Chaos Solitons Fractals*, **13**, 341 (2002).
120. H. G. González-Hernández, J. Alvarez-Gallegos, and J. Alvarez. Experimental analysis of chaos in under-actuated electromechanical systems. *Revista Mexicana de Fisica*, **47**(5), 397 (2001).
121. H. G. González-Hernández, J. Alvarez, and J. Alvarez-Gallegos. Experimental analysis and control of a chaotic pendubot. *Int. J. Robot. Res*., **23(9)**, 891 (2004).
122. J. Alvarez, J. Alvarez-Gallegos, and H. G. Gonz'alez-Hern'andez. Stabilization of unstable periodic orbits for chaotic systems with fractal dimension close to integer. European Control Conference ECC'99, Karlsruhe, Germany, August 31–September 3 (1999).
123. R. S. Mackay, J. D. Meiss, and I. C. Percival. Stochasticity and transport in Hamiltonian systems. *Phys. Rev. Lett*., **52**, 697 (1984).
124. E. E. N. Macau. Targeting in chaotic scattering. *Phys. Rev. E*, **57**, 5337 (1998).
125. Y. C. Lai, T. Te'l, and C. Grebogi. Stabilizing chaotic-scattering trajectories using control. *Phys. Rev. E*, **48**, 709 (1993).
126. Y. C. Lai, M. Ding, and C. Grebogi. Controlling Hamiltonian chaos. *Phys. Rev. E*, **47**, 86 (1993).
127. E. Macau. Using chaos to guide a spacecraft to the Moon. *Acta Astronaut*., **47**, 871 (2000).
128. E. Macau. Exploiting unstable periodic orbits of a chaotic invariant set for spacecraft control. *Celestial Mech. Dyn. Astron*., **87**(3), 291 (2003).
129. M. T. Arjmand, H. Sadeghian, H. Salarieh, and A. Alasty. Chaos control in AFM systems using nonlinear delayed feedback via sliding mode control. *Nonlinear Anal.: Hybrid Syst*., **2**(3), 993 (2008).
130. F. Jamitzky, M. Stark, W. Bunk, W. M. Heckl, and R. W. Stark. Chaos in dynamic atomic force microscopy. *Nanotechnology*, **17**, 213 (2006).
131. M. Ashhab, M. V. Salapaka, M. Dahleh, and I. Mezic. Melnikov-based dynamical analysis of microcantilevers in scanning probe microscopy. *Nonlinear Dyn*., **20**(3), 197 (1999).
132. M. Basso, L. Giarrk, M. Dahleh, and I. Mezic. Complex dynamics in a harmonically excited Lennard–Jones oscillator: micro-cantilever-sample interaction in scanning probe microscopes. *J. Dyn. Syst. Meas. Control*, **122**, 240 (2000).
133. S. Misra, H. Dankowicz, and M. R. Paul. Event-driven feedback tracking and control of tapping-mode atomic force microscopy. *Proc. R. Soc. A*, **464**, 2113 (2008).
134. H. J. Cao, X. B. Chi, and G. R. Chen. Suppressing or inducing chaos by weak resonant excitations in an externally-forced froude pendulum. *Int. J. Bifurcat. Chaos*, **14**(3), 1115 (2004).
135. S. Lenci and G. Rega. Controlling nonlinear dynamics in a two-well impact system. I. Attractors and bifurcation scenario under symmetric excitations. II. Attractors and bifurcation scenario under unsymmetric optimal excitations. *Int. J. Bifurcat. Chaos*, **8**(12), 2387 (1998).
136. S. Lenci and G. Rega. Optimal numerical control of single-well to cross-well chaos transition in mechanical systems. *Chaos Solitons Fractals*, **15**(1), 173 (2003).
137. S. Lenci and G. Rega. Optimal control of homoclinic bifurcation: theoretical treatment and practical reduction of safe basin erosion in the Helmholtz oscillator. *J. Vibr. Control*, **9**(3), 281 (2003).
138. S. Lenci and G. Rega. Optimal control of nonregular dynamics in a Duffing oscillator. *Nonlinear Dyn*., **33**(1), 71 (2003).

139. S. Lenci and G. Rega. Global optimal control and system-dependent solutions in the hardening Helmholtz–Duffing oscillator. *Chaos Solitons Fractals*, **21**(5), 1031 (2004).
140. S. Lenci and G. Rega. Heteroclinic bifurcations and optimal control in the nonlinear rocking dynamics of generic and slender rigid blocks. *Int. J. Bifurcat. Chaos*, **15**(6), 1901 (2005).
141. S. Lenci and G. Rega. Optimal control and anti-control of the nonlinear dynamics of a rigid block. *Phil. Trans. R. Soc. A*, **364**, 2353 (2006).
142. S. Lenci and G. Rega. Control of pull-in dynamics in a nonlinear thermoelastic electrically actuated microbeam. *J. Micromech. Microeng*., **16**, 390 (2006).
143. S. Lenci and G. Rega. Control of the homoclinic bifurcation in buckled beams: infinite-dimensional vs reduced-order modeling. *Int. J. Non-Linear Mech*., **43**(6), 474 (2008).
144. G. Rega and S. Lenci. Identifying, evaluating, and controlling dynamical integrity measures in nonlinear mechanical oscillators. *Nonlinear Anal*., **63**, 902 (2005).
145. G. Rega and S. Lenci. Non-smooth dynamics, bifurcation and control in an impact system. *Syst. Anal. Model. Simul*., **43**(3), 343 (2003).
146. G. Rega and S. Lenci. Dynamical integrity and control of nonlinear mechanical oscillators. *J. Vibr. Control*, **14**, 159 (2008).
147. M. S. Soliman and J. M. T. Thompson. Integrity measures quantifying the erosion of smooth and fractal basins of attraction. *J. Sound Vibr*., **135**, 453 (1989).
148. S. Wiggins. *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. Springer-Verlag, New York, Heidelberg, Berlin (1990).
149. Y. Ketema. A physical interpretation of Melnikov's method. *Int. J. Bifurcat. Chaos*, **2**(1), 1 (1992).
150. S. Lenci and G. Rega. Competing dynamic solutions in a parametrically excited pendulum: attractor robustness and basin integrity.*ASME J. Comput. Nonlinear Dyn*., **3**, 041010-1-9 (2008).
151. J. M. T. Thompson, R. C. T. Rainey, and M. S. Soliman. Mechanics of ship capsize under direct and parametric wave excitation. *Phil. Trans. R. Soc. Lond. A*, **338**, 471 (1992).
152. A. Katz and E. H. Dowell. From single well chaos to cross well chaos: a detailed explanation in terms of manifold intersections. *Int. J. Bifurcat. Chaos*, **4**, 933 (1994).
153. J. M. T. Thompson. Designing against capsize in beam seas: recent advances and new insights. *Appl. Mech. Rev*., **50**, 307 (1997).

Detection of Patterns Within Randomness

Ruedi Stoop and Markus Christen

Abstract The identification of jittered regular signals (="patterns") embedded in a noisy background is an important and difficult task, particularly in the neurosciences. Traditional methods generally fail to capture such signals. Staircase-like structures in the log–log correlation plot, however, are reliable indicators of such signal components. We provide a number of applications of this method and derive an analytic relationship between the length of the pattern n and the maximal number of steps $s(n, m)$ that are observable at a chosen embedding dimension m. For integer linearly independent patterns and small jitter and noise, the length of the embedded pattern can be calculated from the number of steps. The method is demonstrated to have a huge potential for experimental applications.

1 Introduction and Overview

After many years of developing the theoretical fundamentals, dynamical systems have recently started to come up with a rich range of applications, notably in biological systems [1–3]. Characteristic signals of these systems often consist of a mix of regular and noisy signal components. Signals that the neurosciences focus on are the so-called neuronal spike trains, where a neuronal firing event is reduced to a time marker. Since any signal can be characterized by temporal markers, the situation we deal with is a very general one, including, e.g., return-times of dynamical systems and similar observables. Event-marker signals that will serve as illustrative examples in the further development are displayed in Fig. 1.

It has been speculated that spike trains from nervous systems, in particular the human brain, fall into this class of mixed regular-noisy signals [4–14]. As a main generation principle, they are believed to originate from neurons that are mostly driven by complex processes, but occasionally get recruited by more locally defined,

R. Stoop (✉)
Institute of Neuroinformatics, University of Zurich and Swiss Federal Institute of Technology of Zurich ETHZ, CH-8057 Zurich, Switzerland
e-mail: ruedi@ini.phys.ethz.ch

M. Thiel et al. (eds.), *Nonlinear Dynamics and Chaos: Advances and Perspectives,* Understanding Complex Systems, DOI 10.1007/978-3-642-04629-2_12,

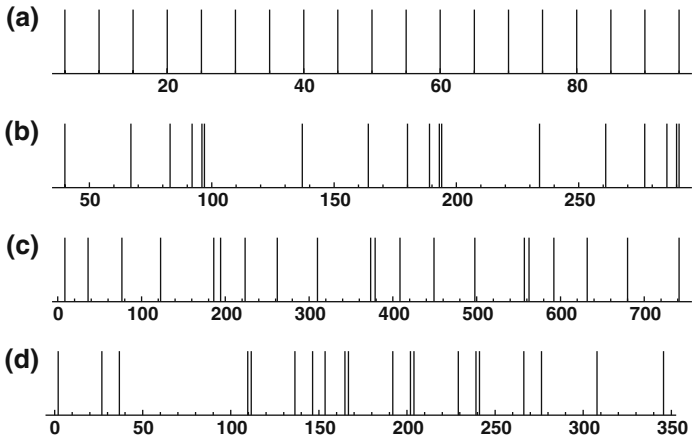


Fig. 1 Temporal event marker signals. (**a**) Length-1 pattern (no jitter). (**b**) Length-5 pattern {40,27,16,9,4,1} (no jitter; pattern will be the guiding example in the main theorem's proof). (**c**) Length-6 pattern {5,24,37,44,59} (with random jitter uniformly drawn from $[-6.4, 6.4]$; pattern is used for Fig. 5d. (**d**) Length-3 pattern {2,25,10} (no jitter, with 25% of the intervals contributed by random spikes uniformly drawn from [0,100]; pattern is used for Fig. 2b

simpler circuits of regular firing. Similar signal characteristics may result from neuronal multi-electrode recordings, where signals from different (randomly, regularly, or in a mixed mode firing) sources arrive at an electrode. After spike sorting by which the events of relevance on the time axis are defined, we may be left with regular signal components embedded in a noisy background.

Both parts of such signals could play equally important roles in cortical signal processing and computation [15]. Here, however, we will be primarily interested in the regular components and how they can reliably be extracted from the data. Past approaches dealing with this task contained several critical tuning parameters, expressing expectations of how the pattern to be searched for should look like, rendering it difficult to assess the validity of the obtained results. This may be one of the reasons why mostly patterns of relatively short length (1–5) have been identified [7, 10], although it may rightfully be argued that information-bearing signals in neuroscience cannot be too long, as actions on relatively short time scales are generally required. In [16], a method for a fast and unbiased detection of patterns in noisy contexts was proposed that does not share these shortcomings. The method is based upon the observation that in the presence of patterns, in the correlation integral plots used for the evaluation of fractal dimensions [17–21], step-like structures emerge. The method works with very modest data size of the kind obtainable in most experimental contexts in neuroscience, and has, in principle, no pattern length limitation. In the presence of a noisy signal component or jitter, the traditional Fourier method, e.g., quickly fails, whereas the characteristic decrease of the number of steps with the embedding dimension is conserved, even for strong noise components. As the only difference from the noise-free case, the most prominent step reappears in a weakened form at multiples of the pattern length, a phenomenon that is simple

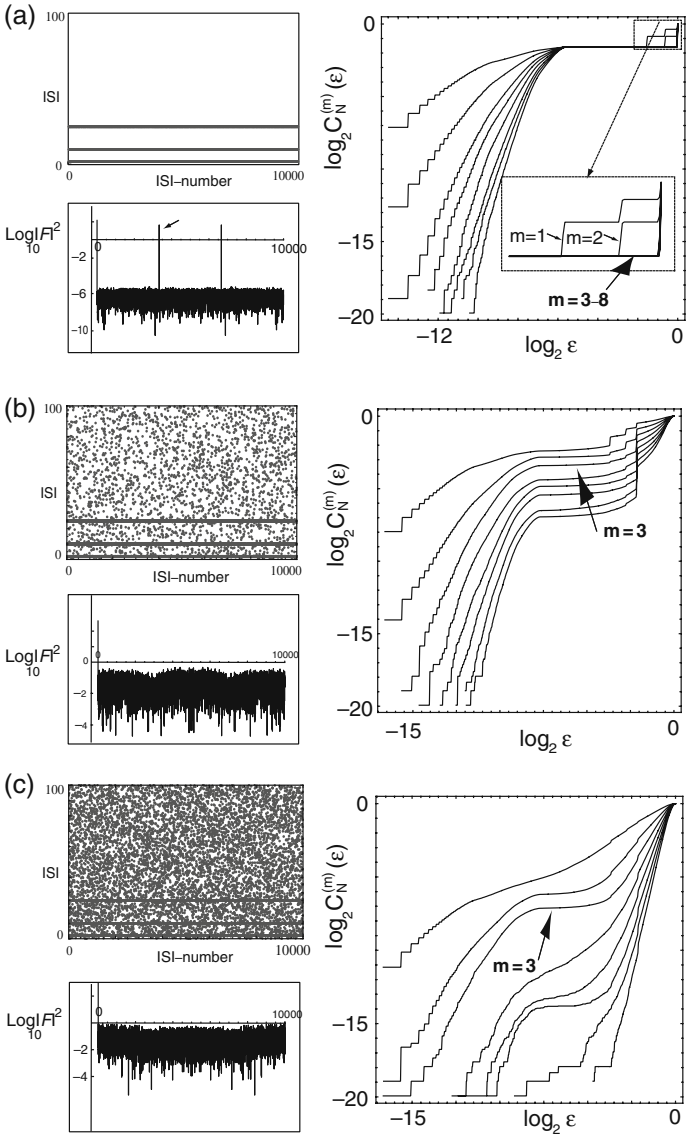


Fig. 2 Comparison of the proposed method [16] with the Fourier approach, using as a paradigm the pattern {2,25,10} with jitter of ±1 % of the smallest interval. Counter-clockwise: ISI-pattern, power spectrum, log–log plot. (**a**) Time series constructed of the pattern only: Both methods predict correctly a pattern of length 3 (*arrows*). (**b**) Time series composed of (whole) patterns and from intervals uniformly randomly distributed in [0,100], so that 25% of all intervals are random: Fourier fails, whereas the log–log plot shows the characteristic decrease of steps with on large step remaining at dimensions 3–8 (*arrow*). (**c**) Same as (**b**) with now 75% of the spike originating from the random process. Fourier fails, whereas the clearest step at dimension 3 indicates a pattern of length 3 (*arrow*). See also Figs. 3 and 7

to understand along the proof of the main theorem given below. The comparison between our method and the traditional Fourier (power spectrum) approach provided in Fig. 2, illustrates these facts.

The derivation of an analytical relationship between pattern length, embedding dimension and number of steps observed, is the main content of this contribution. To make our contribution self-contained, we will proceed as follows. First, we will review the log–log correlation integral method, then point out how the presence of patterns leads to steps in the log–log correlation integral plot, before we arrive at the focus of the present paper, the precise relationship between the length of a pattern embedded in a noisy environment, and the number of maximally identifiable log–log steps. For the analytical derivation of the optimally observable number of steps, we will start with patterns of integer linear independence and absence of noise. Because such experimental conditions are generally not met even on a qualitative level, the relationship is formulated in terms of an upper bound on the number of observable steps.

2 Log–log Steps in the Correlation Integral

Consider an arbitrary scalar time series of measurements $\{x_i\}$, $i = 1 \ldots L$. From this data, embedded points $\xi_k^{(m)}$ are constructed as

$$\xi_k^{(m)} = \{x_k, x_{k+1}, \ldots, x_{k+(m-1)}\}, \tag{1}$$

where m is called the embedding dimension [22, 23]. This *coordinate-delay construction* is standard in nonlinear dynamics [19, 20]. Its purpose is to reconstruct the complete underlying (in general: high-dimensional) dynamics from partial, usually scalar, measurements. The reconstruction of the phase space is generically successful if for a sufficiently large data set a sufficiently high embedding dimension is chosen [19, 20]. Using the embedded points, the *correlation integral* [17, 19–21] is calculated as

$$C_N^{(m)}(\varepsilon) = \frac{1}{N(N-1)} \sum_{i \neq j} \theta\left(\varepsilon - \left\|\xi_i^{(m)} - \xi_j^{(m)}\right\|\right), \tag{2}$$

where $\theta(x)$ is the Heavyside function ($\theta(x) = 0$ for $x \leq 0$ and $\theta(x) = 1$ for $x > 0$) and N is the number of embedded points ($N \leq L - m + 1$). The correlation integral $C_N^{(m)}(\varepsilon)$ averages the probability of measuring a distance smaller than ε between two randomly chosen points $\xi_i^{(m)}$ and $\xi_j^{(m)}$. In practical applications, $\log C_N^{(m)}(\varepsilon)$ is plotted against $\log \varepsilon$ (the so-called *log–log plot*). The correlation dimension $d_C^{(m)}$ is defined as the limit $d_C^{(m)} = \lim_{\varepsilon \to 0} \frac{\log C_N^{(m)}(\varepsilon)}{\log \varepsilon}$ [17, 19, 20]. If an embedding dimension $m > 2d_C^{(m)}$ is chosen, the slope of $\log C_N^{(m)}(\varepsilon)$ versus $\log \varepsilon$ for small ε provides a good estimate of the correlation dimension.

For the evaluation of the distances, any norm could be used. Instead of the "natural" Euclidean norm, we will mainly use the maximum norm, in order to simplify numerical computations and theoretical arguments. Degeneracies introduced by this choice are removed upon the addition of a small amount of noise. For an approach extracting fractal dimensions from noisy time series (which is not one of our goals), see [24].

We first demonstrate how the presence of patterns leads to a step-like behavior of the log–log correlation dimension plots. Patterns manifest themselves as a clustering of the embedded data. For the calculation of $C_N^{(m)}(\varepsilon)$, an embedded point $\xi_0^{(m)}$ is chosen at random. As the radius ε of its neighborhood $U\left(\xi_0^{(m)},\varepsilon\right)$ is enlarged, we keep track of the number of points that fall into this neighborhood. If a point newly entering the neighborhood belongs to a cluster, upon a small enlargement of ε, many points will join. That is, the number of points $C_N^{(m)}(\varepsilon)$ quickly increases with ε. Once the cluster size is reached, fewer points are recruited, and $C_N^{(m)}(\varepsilon)$ increases but slowly. In this way, step-like structures emerge. The denser the clustering regions, the more prominent the step-like structure. To demonstrate this effect, artificial (noise-free) time series were constructed from a repetition of a sequence of length n. The series were then embedded (using embedding dimension m) and the correlation integrals were evaluated. The results shown in Fig. 3 demonstrate a clean emergence of stairs, the number of which increases with the length of the embedded pattern n, and decreases with the embedding dimension m. In the presence of patterns, the step-like behavior emerges stably even for a few hundred scalar measurements. If in an experiment single trials generate less data than needed (say, in neuroscience, because of adaptation), data from several trials under identical conditions can be concatenated. Although in this case the embedded data will contain some points that violate the continuous dependence on time, this has normally no statistical influence.

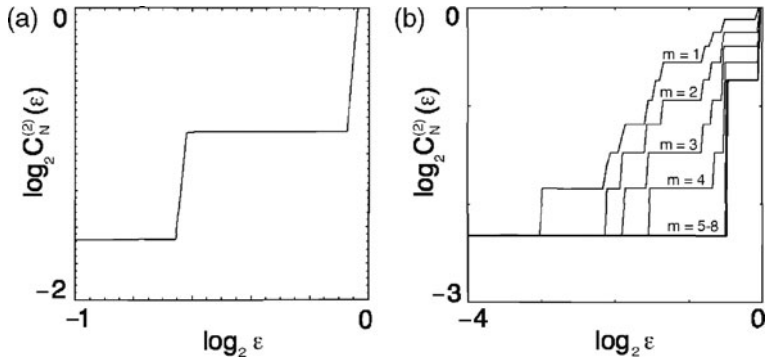


Fig. 3 Correlation integral of time series generated by a repeated sequence of length n. (**a**) Emergence of steps in the log–log plot of $C_N^{(m)}(\varepsilon)$) ($m = 2$, $n = 3$, pattern $\{1, 2, 4\}$, maximum norm ε rescaled). (**b**) Decrease of the number of steps with increasing embedding dimension m ($m = 1, ..., 8$, $n = 5$)

The basis of the relationship between number of steps s, pattern length n and embedding dimensions m is provided by the following Proposition.

Proposition 1 *The number of correlation-dimension log–log steps $s(n,m)$ generated from an embedded repeated pattern, equals the number of distinct distances among the embedded points using the maximum norm.*

Proof For the correlation integral, all $\frac{n(n-1)}{2}$ distances between points are calculated, where classes of equal distances $\{\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_\kappa\}$ are generated. Locally, around a point $\xi_0^{(m)}$ in the embedding space, $C_N^{(m)}(\varepsilon,x_k)$ changes whenever $\varepsilon \in \{\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_\kappa\}$. As this is true for any point, also for the averaged correlation integral $C_N^{(m)}\left(\varepsilon,\xi_0^{(m)}\right)$ the number of steps is $s(n,m) = \kappa$. This proves the proposition. □

3 Noiseless Single Patterns and Beyond

After having shown that the presence of patterns is reflected in the emergence of log–log correlation integral steps, our next goal is an estimate of the pattern length from the number of steps. That this might be achievable is motivated by the following argument. Using the maximum norm, the distance between two points in the embedding space is defined as the maximum of the component differences. As the dimensionality of the embedding space is increased, ever more of the possible differences will be present. A few large differences will, however, prevent the smaller ones from winning the competition for the maximum. As a consequence, the number of steps $s(n,m)$ obtained for a pattern of length n can be expected to decrease with increased embedding dimension m. That this indeed is the case is demonstrated in Fig. 3b.

The precise way how this decay proceeds depends on the pattern length n. For toy systems, the maximal number of occurring steps $s(n,m)$ can be computed numerically as follows. A time series generated by repeating a sequence of length n composed of elements $\{x_1,\ldots,x_n\}$, generates distinct coordinate differences $d_{ij} := |x_i - x_j|$. By shifting a window of length m along the time series, we repeatedly generate embedded points of embedding dimension m. On the set of the generated points, the maximum norm induces classes of equal distances, the number of which equals $s(n,m)$. Unfortunately, this numerical calculation quickly exhausts computing time, calling for an analytical way to compute $s(n,m)$. The values of $s(n,m)$ that can be corroborated with the help of a desktop computer are shown in Table 1. In Fig. 7 we demonstrate how for the toy system generated from the sequence $\{5, 24, 37\,44, 59\}$, the correlation integral method is able to reproduce the decrease of $s(n,m)$ predicted by Table 1: In embedding dimension $m = 1$, all ten possible differences are detected. As m increases towards 5, the number of steps decreases in accordance with Table 1, before remaining constant for $m > 5$.

In experimental applications, one might expect more than one pattern to be present in a time series. This leads to complications in the application of $s(n,m)$. If, for the simplest case, one single step emerges in the log–log plot, this could either be

Table 1 $s(n, m)$ for $n, m = 1, \ldots, 10$

m\n	1	2	3	4	5	6	7	8	9	10
1	0	1	3	6	10	15	21	28	36	45
2	0	1	2	4	8	12	18	24	32	40
3	0	1	1	3	6	9	15	20	28	35
4	0	1	1	2	4	7	12	16	24	30
5	0	1	1	2	2	5	9	13	20	25
6	0	1	1	2	2	3	6	10	16	21
7	0	1	1	2	2	3	3	7	12	17
8	0	1	1	2	2	3	3	4	8	13
9	0	1	1	2	2	3	3	4	4	9
10	0	1	1	2	2	3	3	4	4	5

due to one pattern composed of length two, or to two "patterns" of length one each. A greater number of steps, as obtained from a multitude of patterns, will further complicate this problem. Once the presence of patterns is indicated and precise alternatives for the patterns are posed by the method, the existence/non-existence of a particular alternative, can be corroborated by direct methods that under such conditions are justified. The number of steps predicted by $s(n, m)$ is reliable up to very strong noise components (this will be demonstrated in Fig. 7) or up to the point where the jitter of the pattern conflicts with its nature. Even in very difficult conditions, the method is able to indicate the presence of patterns, where $s(n,m)$ can still serve as a guideline for further processing (see the final discussion).

4 Analytical Derivation of *s(n, m)*

For the analytical derivation of $s(n, m)$ we start from a time series $\{x_i\}_{i=1\ldots N}$ generated by the repetition of a pattern of length n. The pattern is supposed to be general in the sense that the elements $\{x_1, \ldots, x_n\}$ yield $n(n-1)/2$ distinct coordinate differences $d_{ij} = |x_i - x_j|$ (="integer linear independence")). Embedded points are generated by the shift of a window of length m along the data series $\{x_i\}_{i=1\ldots N}$. As was previously pointed out, on the set of the embedded points the maximum norm induces classes of equal distances, the number of which equals $s(n, m)$. The goal of the rest of this paper is to analytically compute $s(n, m)$.

We start by calculating the number of different distance vectors. Since different distance vectors will not necessarily imply different distances between points, this is but a preliminary task that can be achieved without specifying the metric. Using this information, we will calculate the number of different distances, specifying the maximum norm as the relevant one.

We will focus on the case $m \leq n$ (case $m > n$ is trivial). As an example of $n = 4, m = 3$, we start with the repeated sequence $\{x_1, x_2, x_3, x_4\}$, which generates the time series

$$\{x_1, x_2, x_3, x_4, x_1, x_2, x_3, x_4, x_1, x_2, x_3, x_4, x_1, x_2, ..\}. \tag{3}$$

By the embedding process in $m = 3$, we obtain the set of embedded points

$$\{\{x_1, x_2, x_3\}, \{x_2, x_3, x_4\}, \{x_3, x_4, x_1\}, \{x_4, x_1, x_2\}\}.$$

The distance vector components d_{ij} between the embedded points then become

$$\begin{aligned} |\{x_1, x_2, x_3\} - \{x_2, x_3, x_4\}| &= |\{x_1 - x_2, ., x_3 - x_4\}| =: \{d_{12}, d_{23}, d_{34}\}, \\ |\{x_2, x_3, x_4\} - \{x_3, x_4, x_1\}| &= |\{x_2 - x_3, ., x_4 - x_1\}| =: \{d_{23}, d_{34}, d_{41}\}, \\ |\{x_3, x_4, x_1\} - \{x_4, x_1, x_2\}| &= |\{x_3 - x_4, ., x_1 - x_2\}| =: \{d_{34}, d_{41}, d_{12}\}, \end{aligned} \tag{4}$$

where $|\ .\ |$ indicates the componentwise 'absolute value' operation. The emerging distance vectors can be collected in the form of a 2-torus:

$$D_{(n)} = \left|\begin{array}{cccccc} d_{12} & d_{13} & \cdots & \cdots & \cdots & d_{1n} \\ d_{23} & d_{24} & \cdots & \cdots & d_{2n} & d_{21} \\ \vdots & \cdots & \cdots & \cdots & \cdots & \vdots \\ d_{(n-1)n} & d_{(n-1)1} & \cdots & \cdots & \cdots & d_{(n-1)(n-2)} \\ d_{n1} & d_{n2} & \cdots & \cdots & d_{n(n-2)} & d_{n(n-1)} \\ \hline d_{12} & d_{13} & \cdots & \cdots & \cdots & d_{1n} \\ d_{23} & d_{24} & \cdots & \cdots & d_{2n} & d_{21} \\ \vdots & \cdots & \cdots & \cdots & \cdots & \vdots \\ d_{(n-1)n} & d_{(n-1)1} & d_{(n-1)2} & \cdots & d_{(n-1)(n-3)} & d_{(n-1)(n-2)} \end{array}\right| . \tag{5}$$

In the distance matrix, distance vectors in an m-dimensional embedding space are represented by sub-columns of dimension m. As an illustration, for $n = 5$ we obtain

$$D_{(5)} = \left|\begin{array}{cccc} d_{12} & d_{13} & d_{14} & d_{15} \\ d_{23} & d_{24} & d_{25} & \mathbf{d_{21}} \\ d_{34} & d_{35} & \mathbf{d_{31}} & d_{32} \\ d_{45} & \mathbf{d_{41}} & d_{42} & d_{43} \\ \mathbf{d_{51}} & d_{52} & d_{53} & d_{54} \\ \hline d_{12} & d_{13} & d_{14} & d_{15} \\ d_{23} & d_{24} & d_{25} & d_{21} \\ d_{34} & d_{35} & d_{31} & d_{32} \\ d_{45} & d_{41} & d_{42} & d_{43} \end{array}\right| .$$

Because of the two-torus nature of D, it is no surprise that by starting the columns at arbitrary positions we observe the occurrence of repeated sub-columns (above, locations leading to repetitions are shown in blue), even though we have requested that $d_{ij} \neq d_{kl}$, unless $_{ij} =_{kl}$ or $_{kl} =_{ji}$.

We will call the upper part of $D_{(n)}$ the *top rectangle* and the lower part of $D_{(n)}$ the *bottom rectangle*. In the top rectangle, the triangle defined by the edges d_{12}, d_{1n} and

$d_{(n-1)(n)}$ will be called the *left triangle*; the triangle defined by the edges d_{n1}, d_{21} and $d_{n(n-1)}$ will be called the *right triangle* (see Fig. 4). It is worthwhile keeping in mind that the indices $_{ij}$ of d_{ij} do *not* reflect the position of d_{ij}.

Using these notations, we will prove the following proposition:

Proposition 2 *The set of distinct sub-columns of dimension m ($m \leq n$ arbitrary but fixed) of the difference matrix, is equivalent to the set of sub-columns starting in the left triangle of the difference matrix. Consequently, their number is given by*

$$\Omega := \frac{n(n-1)}{2}. \tag{6}$$

Proof Sub-columns starting in the left triangle are distinct. We will show that all sub-columns that start in the right triangle or in the bottom rectangle, are equivalent to a sub-column starting in the left triangle. We split the proof into three steps. Due to the construction of the difference matrix, it is obvious that all sub-columns that lie completely inside the bottom rectangle have a counterpart in the top rectangle. Sub-columns starting in the bottom rectangle are therefore repeating sub-columns. Moreover, since $d_{ij} = d_{ji}$, all sub-columns that lie completely inside the right triangle have a counterpart in the left triangle and therefore are repeating sub-columns.

For sub-columns that cross the border between the rectangles and/or between the triangles it suffices to show that all sub-columns starting in the right triangle and crossing the border between top and bottom rectangle, have a counterpart that starts in the left triangle. To this end we consider the borders between rectangles and triangles (border T and border B, and border L and border R, respectively, see Fig. 4). Since $d_{ij} = d_{ji}$, the elements of border R $(d_{n,1}, \ldots, d_{2,1})$ reappear at the border B $(d_{1,2}, \ldots, d_{1,n})$, in reversed order. Also the elements of border L $(d_{n-1,n}, \ldots, d_{1,n})$ appear at the border T $(d_{n,1}, \ldots, d_{n,n-1})$, also in reversed order. Sub-columns starting in the right triangle and crossing from T to B have always a counterpart sub-column starting in the left triangle and crossing from L to R.

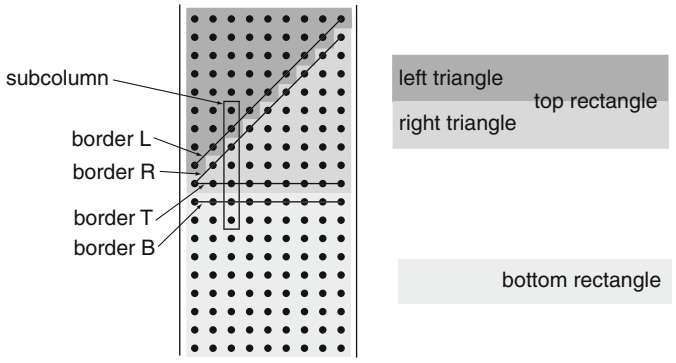


Fig. 4 Partition of the difference matrix (see text)

For all sub-columns that either start in the right or in the bottom rectangle, there is therefore a sub-column starting in the left triangle. Consequently, the set of distinct sub-columns can be obtained by counting the sub-columns of the left triangle, the number of which evaluates to $\frac{n(n-1)}{2}$. This proves the proposition. □

It has already been pointed out that distinct sub-columns do not necessarily imply distinct point distances. To count the number of distinct distances we will use the maximum norm, which considerably simplifies the arguments. Evaluation of the maximum norm amounts to the determination of the largest element from the sub-column. As the differences d_{ij} are all distinct, they can be ordered according to size, with $\bar{d}_{ij}$ denoting the largest element. For $m \geq 2$, $\bar{d}_{ij}$ will appear in several distinct sub-columns that all lead to the same distance. Moreover, the explicit decrease of $s(n,m)$, for n fixed and m increasing, depends on the specific order of the d_{ij} in the difference matrix. Therefore, only an upper bound $s(n,m)$ of the number of distinct distances can be given. To determine this bound we consider an order of d_{ij} that maximizes the number of observable steps.

Proposition 3 *The number of distinct distances between the embedded points is maximal if the elements of the columns in the left triangle are monotonously decreasing (increasing)*

$$\left|\begin{array}{cccccc} d_{12} & d_{13} & \cdots & d_{1(n-1)} & d_{1n} \\ \vee & \vee & \cdots & \vee & \\ d_{23} & d_{24} & \cdots & d_{2n} & \\ \vee & \vee & \cdots & & \\ \vdots & \vdots & \vdots & & \\ \vee & \vee & & & \\ d_{(n-2)(n-1)} & d_{(n-2)n} & & & \\ \vee & & & & \\ d_{(n-1)n} & & & & \end{array}\right| . \tag{7}$$

Proof It suffices to prove Proposition 3 for monotonously decreasing distance elements. First we will show that there exists a sequence $(a_1,\ldots,a_n)$ with the required ordering of the differences. Then we will show that this ordering leads to a maximal number of distinct distances.

We proceed by construction: For the series $a_{n-\iota} = 2^\iota$ for $\iota = l,\ldots,n-1$, we obtain distinct nonzero differences $d_{ij} = |a_i - a_j|$. For two consecutive elements in the l-th column of the left triangle, we obtain

$$\begin{aligned} d_{n-(\iota+1),n-(\iota+1-l)} - d_{n-\iota,n-(\iota-l)} = \\ |a_{n-(\iota+1)} - a_{n-(\iota+1-l)}| - |a_{n-\iota} - a_{n-(\iota-l)}| = \\ |(2^{\iota+1}-1) - (2^{\iota+1-l}-1)| - |(2^\iota - 1) - (2^{\iota-l}-1)| = \end{aligned} \tag{8}$$

$$\sum_{p=\iota-l+1}^{\iota} 2^p - \sum_{p=\iota-l}^{\iota-1} 2^p = 2^\iota - 2^{\iota-l} > 0.$$

The constructed sequence therefore has the desired ordering within column l. By construction, the elements d_{ij} of the difference matrix are distinct and nonzero, so that they can be ordered according to size. We first focus on sub-columns that lie *entirely* within the left triangle. We will show that for a maximal number of distinct point distances, $\bar{d}_{ij}$ of each column must be located in the first row.

Consider in the left triangle a column l and assume that $\bar{d}_{ij}$ is in row i, with $1 < i \leq n - l$. Then, for $2 \leq m \leq i$, there are up to m sub-columns starting in row $i-m+1$ and lying completely in the left triangle, that include $\bar{d}_{ij}$. As a consequence, they will generate the identical (maximal) distance. Only when $\bar{d}_{ij}$ is located in the first position of column l, it will be in precisely one sub-column, and the number of distinct distances will be maximal, see Fig. 5a. In order to generate a maximal number of distances, the elements d_{ij} must therefore have the requested ordering.

We now consider sub-columns that start in the left triangle but extend to the right triangle and possibly also to the bottom rectangle (see Fig. 4). Again, the number of distinct distances would be maximal, if *all* elements in *every* row of the difference matrix would take over the ordering of the top triangle. Unfortunately, this is not possible, since there are constraints among the elements of the matrix. However, the desired ordering of the elements d_{ij} can also be achieved in the columns of the right triangle, since $d_{ij} = d_{ji}$. Moreover, if we only consider sub-columns starting in the left triangle with dimension is not larger than m, the desired ordering is also be achieved by the sub-column elements of the bottom rectangle, since in this part of the bottom rectangle, the left triangle is repeated.

The obstruction against the desired monotonous decrease of d_{ij} emerges at the borders $\{T, B\}$ and $\{L, R\}$. We will show that despite these potential disruptions of the ordering, the number of distinct distances between the embedded points will be maximal. We deal with the two borders separately.

For reasons of simplicity, we first deal with the $\{T, B\}$-*border*, considering later the impact of an additional disruption of the monotonous ordering at the $\{L, R\}$-border. By construction of the matrix, the biggest element $\bar{d}_{ij}$ of a column in the left triangle will show up again in the bottom rectangle, so that in each column at some point, the monotonous decrease of the values d_{ij} will be disrupted. If the order of the elements d_{ij} is according to the Proposition 3, the disruption happens at the $\{T, B\}$-border.

We argue that despite this disruption, the ordering to the Proposition 3 will be optimal. Consider a column l of the left triangle with largest element $\bar{d}_{ij}$. This element will repeat in the bottom rectangle. Independently of the position of $\bar{d}_{ij}$ in the left triangle, the number of intermediate d_{ij} between the two locations $\bar{d}_{ij}$ is $n - 1$. An element $\bar{d}_{ij}$ in the bottom rectangle can therefore only be part of a sub-column, if the latter's dimension m is large enough (except for the first few columns of the

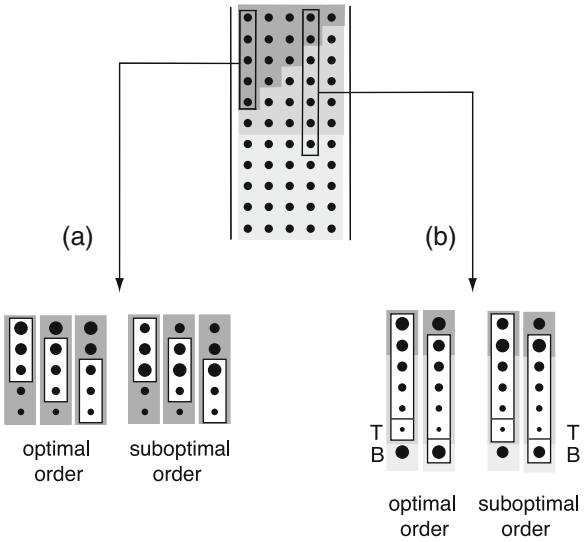


Fig. 5 Illustration of the proof of Proposition 3 (pointsize represents size of d_{ij}): (**a**) If the ordering of the elements of the difference matrix is according to the proposition, each location of a sub-column of dimension 3 results in a new distance. If, however, the largest element is placed in the third place, all sub-columns result in the same distance. (**b**) The disruption of the monotonous decrease at the border $\{T, B\}$ only affects sub-columns of a large dimension m (in the given example, for $m = 6$ only one distance each for the optimal and the suboptimal case). For sub-columns with smaller m, the ordering within the *left* triangle is relevant (for $m = 5$ in the given example we obtain two distances in the optimal case, and one in the suboptimal case). The distinction of the cases is similar to the one expressed by k (Eq. (10))

difference matrix). Therefore, the position of $\bar{d}_{ij}$ in the left triangle is crucial in order to maximize the number of distinct distances. As has been shown before, in order to achieve this, $\bar{d}_{ij}$ needs to be at the first position in the column (see Fig. 5b).

A monotonous decrease of d_{ij} beyond the {*L*,*R*}-*border* would require all of the following inequalities to be fulfilled:

$$d_{(n-1)n} > d_{n1}, \quad d_{(n-2)n} > d_{(n-1)1}, \quad \ldots, \quad d_{1n} > d_{21}. \tag{9}$$

Due to the dependencies between the element of the matrix, this cannot be satisfied. Consider for example the first and the last column of the top rectangle. From a monotonous decrease in the last column of the top rectangle it would follow that $d_{1n} > d_{n(n-1)}$, contradicting the first inequality. Continuing in this way by comparing the second column with the second-last column, etc., it becomes obvious that in the chain of inequalities (9), either the first or the second part of the chain can be fulfilled, but not both of them. Since the left triangle sub-columns starting in the first columns contain more {*T*,*B*}-border elements than those starting in the last columns, for obtaining a maximal number of distinct distances, the fulfillment

of the second part of (9) should be aimed at. This is achieved by the proposed ordering.

The distinction between the two cases is given by the row number k such that the top sub-rectangle containing the first k rows is monotonously ordered. k is evaluated as

$$k = \begin{cases} n/2 & : \quad \text{even } n \\ (n-1)/2 & : \quad \text{odd } n \ . \end{cases} \tag{10}$$

This proves the proposition. □

As an example consider the repeated pattern generated from the sequence (40,27, 16,9,4,1) ($n = 6$) with distance matrix

$$\left|\begin{array}{ccccc} \mathbf{13} & \mathbf{24} & \mathbf{31} & \mathbf{36} & \mathbf{39} \\ \mathbf{11} & \mathbf{18} & \mathbf{23} & \mathbf{26} & 13 \\ \mathbf{7} & \mathbf{12} & \mathbf{15} & 24 & 11 \\ \mathbf{5} & \mathbf{8} & 31 & 18 & 7 \\ \mathbf{3} & 36 & 23 & 12 & 5 \\ 39 & 26 & 15 & 8 & 3 \\ \hline 13 & 24 & 31 & 36 & 39 \\ 11 & 18 & 23 & 26 & 13 \\ 7 & 12 & 15 & 24 & 11 \\ 5 & 8 & 31 & 18 & 7 \\ 3 & 36 & 23 & 12 & 5 \end{array}\right| \quad .$$

For this pattern, from Eq. (10) we obtain $k = 3$, implying a monotonous behavior in the first three (sub)columns of the left triangle {13,11,7,5,3}, {24,18,12,8}, {31,23,15}, and in the last two columns of the top rectangle {36,26,24,18,12,8}, {39,13,11,7,5,3}. Elements of the left triangle are in bold type.

Proposition 4 *Given an embedded time series generated from a repeated pattern* $(x_1, \ldots, x_n)$. *If the ordering of the associated difference matrix in the left triangle is according to Proposition 3, the number of distinct distances will be maximal and given by*

$$s(n,m) = \Omega - \Lambda + \Pi - \Upsilon + \Gamma, \tag{11}$$

where

$$\Lambda = k(m-1), \tag{12}$$

$$\Pi = \frac{i(i+1)}{2} \quad \textit{where} \quad \begin{array}{l} \textit{n even: } i = m-k-1 \textit{ if } \ m > k+1 \textit{ else } \ i = 0 \\ \textit{n odd: } \ i = m-k-2 \textit{ if } \ m > k+2 \textit{ else } \ i = 0, \end{array}$$

$$\Upsilon = \begin{cases} \frac{i(i+1)}{2} & \textit{if} \quad m > k+2 \quad \textit{and} \quad i = m-k-2 \\ 0 & \textit{else,} \end{cases}$$

$$\Gamma = \begin{cases} 1 \ n\textit{:even} \quad \wedge \quad m > k \\ 0 \ \textit{else.} \end{cases}$$

Proof For the calculation, we start from the number of distinct difference sequences that start in the left triangle $\Omega = \frac{n(n-1)}{2}$ (Eq. (6)), among which, however, there will be some multiply generated distances. If the elements d_{ij} are ordered according to Proposition 3, obvious multiply generated distances are those from $1, \ldots, k$-subcolums crossing the $\{L,R\}$-borders, because in each column of the top rectangle, the R-border elements provide the largest elements. Therefore, most sub-columns starting in the first k columns of the left triangle and containing elements of the R-border yield distances identical to those from sub-columns starting in the columns $k+1, \ldots, n-1$. By the term "most" we indicate, that for n even and $m > k$, a special case needs to be analyzed (to be included in the term Γ, see below). Similarly, for $k+1, \ldots, n-1$, the B-border elements provide the largest components.

Λ: In order to count the number of sub-columns of dimension m containing elements of the left triangle as well as R-border elements, we first also accept subcolumns of dimension m that potentially 'stick out' of the matrix (see Fig. 6a). The number of these d_{ij}-positions covered is $\Lambda := k(m-1)$, with k as in Proposition 3, Eq. (10). The equal number of subcolums (starting at these positions) does not lead to new distances and therefore needs to be subtracted from Ω.

Π: In Λ we also counted a number Π of virtual sub-columns (that "stick out" of the matrix and are therefore of a virtual nature). Their number depends on whether

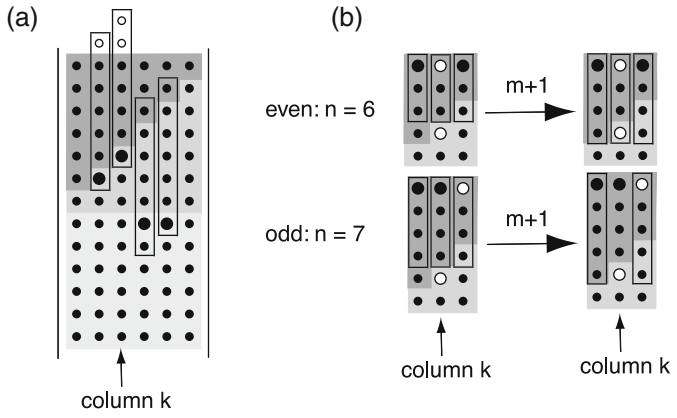


Fig. 6 Proposition 4 illustrated: **(a)** Counting subcolumnns crossing the $\{L,R\}$- and $\{T,B\}$-borders: Λ counts subcolumns until column k that contain elements of the R border, including some "stick-out" subcolumns. Expression Π counts the number of stick-out subcolumns. Circles mark the positions of non-existing subcolumns ($n = m = 7$). Υ counts the subcolumns in column $k+1$ that cross the $\{T,B\}$-border. Larger subcolumn elements are indicated by larger dots. **(b)** Special case accounted for by Γ. For $n = 6$, the subcolumns in the k'th column with dimension $m > 3$ generates a distinct distance, whereas for $n = 7$, the subcolumn of dimension $m > 4$ fails to do so. Open circles: Locations of the two elements of identical largest size. For n even they are in the same column $d_{1(k+1)}$

n is even or odd. One easily derives that Π has the value $\Pi = \frac{i(i+1)}{2}$, where for even n, $i = m - k - 1$ (and $m > k + 1$), and for odd n, $i = m - k - 2$ (and $m > k + 2$). This number has to be added to $\Omega - \Lambda$.

Υ: Now we consider sub-columns starting from the left triangle of columns with numbers $k + 1, \ldots, n - 1$. In these columns, a monotonous decrease of d_{ij} can be achieved until the $\{T,B\}$-border is reached. Therefore, sub-columns starting in columns $l = k + 1, \ldots, n - 1$ of the left triangle containing B-border elements yield a distance already obtained by sub-columns starting from the first row in column l (see Fig. 6a). Thus, their number is $\Upsilon = \frac{i(i+1)}{2}$, where $i = m - k - 2$ and $m > k + 2$. Υ needs to be subtracted from $\Omega - \Lambda + \Pi$.

Γ: Finally, we take into account the special case n even and $m > k$. Due to construction, the largest element in the k-th matrix row is the element $d_{1(k+1)}$, which appears in row number 1 and $k + 1$. Sub-columns starting in the left triangle with $m \geq k + 1$ therefore all yield the same distance (see Fig. 6b). This shortcoming can be compensated for by a term $\Gamma := 1$ for even n and $m > k$ and $\Gamma := 0$ else. Γ has to be added to $\Omega - \Lambda + \Pi - \Upsilon$.

We now have taken into account all sub-columns crossing the $\{L, R\}$- and the $\{T,B\}$-border in the difference matrix, showing which one of them generate already generated distances. This proves the proposition. □

Proposition 5 *For $m > n$, the number of distinct distances equals $s(n,m = n)$ (and is therefore independent from m).*

Proof Consider sub-columns of length $m = n$ starting in the left triangle. Due to the periodicity of the sequence $(a_1, \ldots, a_n)$, for $m > n$, d_{ij}-elements entering a sub-column from below are already contained in the sub-column. Therefore, no new maximal d_{ij} can enter these sub-columns, so that $s(n,m)$ cannot change any more. This proves the proposition. □

5 Main Theorem

Collecting all terms, we arrive at the following result: The maximal number of steps $s(n,m)$ emerging in a log-log plot of a time series from a repeated pattern of length n in embedding space dimension m has the expression

$$n \quad \text{even:} \quad s(n,m) = \begin{cases} \frac{n(n-m)}{2} & : \quad 1 \leq m \leq \frac{n}{2} \\ \frac{n(n-m)+2m-n}{2} & : \quad \frac{n}{2} < m \leq n \\ \frac{n}{2} & : \quad m > n, \end{cases} \tag{13}$$

$$n \quad \text{odd:} \quad s(n,m) = \begin{cases} \frac{n(n-m)+m-1}{2} & : \quad 1 \leq m \leq n \\ \frac{n-1}{2} & : \quad m > n \end{cases} \tag{14}$$

Using this analytic expression, we obtain the table of the maximal number of observable steps $s(n,m)$ as presented in Table 1.

6 Discussion and Outlook

We have investigated how the number of steps s emerging in the log–log correlation integral plot of time series generated from repeated patterns depends on the the length of the pattern n and the embedding dimension m. We were able to derive and prove an analytical relationship between the three quantities. That this relationship continues to hold for mixed noise-pattern signals, for noise components where alternative methods fail or are inconclusive, has been demonstrated in Fig. 2. Up to strong noise components, the derived formula allows for a direct determination of the pattern length (and the associated length scales) if only one integer linearly independent pattern is present in the time series. As in applications these conditions may be violated or difficult to verify (e.g. more than one pattern might be present), the result was formulated as a statement on the maximal number of observable steps.

A final point of interest not discussed so far is the influence of jitter on the repeated patterns. In general, jitter modifies the density of the point clusters in the embedding space and, therefore, the distribution of the distances. In the log-log plot this primarily leads to a smearing of the steps (some consequences thereof have already been investigated in [16]). A pattern will always emerge in the embedded time series in its most genuine form (it is neither cut into pieces, nor spoiled by foreign points) if the embedding dimension equals the pattern length ($n = m$). In the case of absence of a noise component, the decrease of steps fully stops at $n = m$, whereas in the presence of a noise component, the steps repeat at multiples of n, in a softened fashion. As a consequence, the most prominent step can be expected to yield a reliable indicator of the pattern length.

That this is indeed the case is demonstrated in our final example, where to the series generated from the sequence {5,24,37,44,59} (cf. Table 1), jitter was added. For the following, we define the strength of the jitter as the ratio of the interval size from which we uniformly sample the jitter, over the shortest interval of the pattern. The results (Fig. 7b–f) demonstrate that the pattern length can be reliably estimated up to a jitter strength of 512% (Fig. 7e), as the most pronounced step still appears at $m = 5$. The number of steps for $m < 5$, however, are affected by the jitter: For $m = 1$, for example, 9 steps are identifiable at 8% jitter (Fig. 7b), 7 steps at 32% (Fig. 7c) and 3 steps at 128% (Fig. 7d). The step-like structure disappears if the jitter reaches the size of the largest element of the pattern (Fig. 7f). Thus, the criterion that the most pronounced step appears at $m = n$, still yields a valuable indicator for the pattern length at strong jitter.

In experimental applications, one might have more than one pattern in a time series. This complicates the expression for $s(n, m)$. If, for the simplest case, one single step emerges in the log–log correlation integral plot, this could either be due to one pattern composed of length two, or to two “patterns” of length one each.

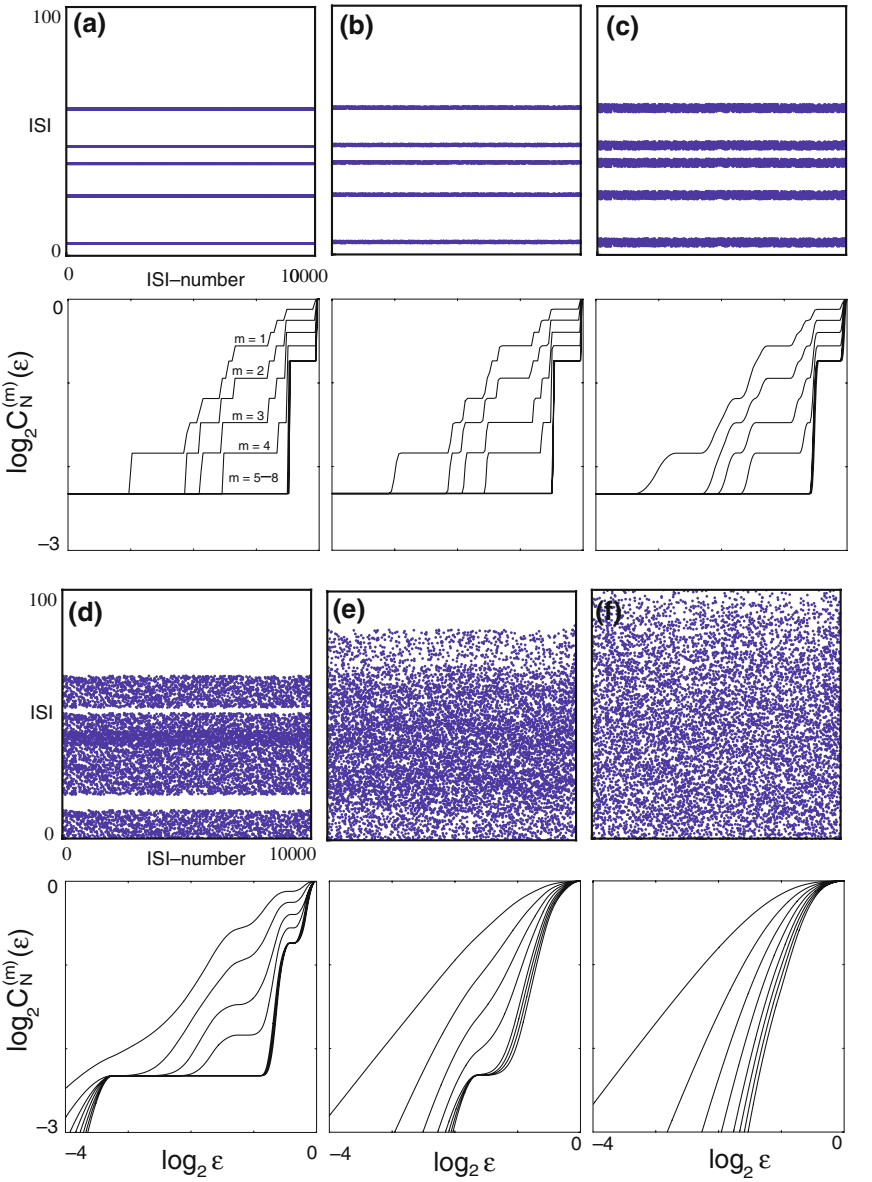


Fig. 7 Step number decrease with increasing embedding dimension, in the presence of background noise ($n = 5, m = 1, \ldots, 8$). *Upper row*: ISI pattern, *lower row*: log-log plots (**a**) For $m = 1$ we find 10 steps, in agreement with Table 1. Panels (**b**)-f): Noise levels 8, 32, 128, 512 and 1,024%. Despite the step number decrease, the clearest step always emerges for $n = 5$, indicating the presence of a pattern of length 5

A greater number of steps, as obtained from a multitude of patterns, further complicates the situation. Fortunately, even in this case, the method is able to provide the correct information on the length of the ingredients for a mixture of linearly independent patterns, notably even in the presence of very general kinds of background noise [16]. For covering these non-optimal cases and conditions, the Main Theorem was formulated as a statement on the maximal number of observable steps for a repeated pattern. It is worth emphasizing that once the presence of patterns is evident and suggestions of patterns have been obtained by our method, their existence/non-existence can be confirmed by direct methods. If based on the insight provided by our method, the direct methods are better justified and more efficient.

To illustrate the potential of the approach, we briefly mention one recent experimental application.[1] In this experiment, motor neuron re-aggregates of spherical extension of 120–150 μm size were cultivated on a 2-D micro-electrode array, where the spacing between electrodes is 200 μm and the electrode diameter is 30 μm. The neurons form a network, of which the electrodes acquire the extracellular potentials fluctuations (Fig. 8a). On the other side of the chip, muscle cells are cultivated. Applying our approach to spike trains derived from extracellular potentials data, we found that during in vitro day 3, at one particular electrode a pattern of length 5 was present (Fig. 8b). At all other investigated electrodes, in contrast, only noisy signals were detected (Fig. 8c). The firing frequency at the pattern generating site was significantly lower than the frequencies at the other electrodes. After day 5, the pattern vanished. These findings suggest that a potential for a pattern generator is initially built up at this site, but then is discarded during further development. Experimental efforts now attempt to clarify the origin of the pattern generator and its conditions of sustainability.

In our contribution, we have unveiled how patterns account for the emergence of steps in Log–log correlation integrals plots. Moreover, we have shown that these steps provide the basis for an efficient and robust identification of patterns present within noise, on all scales of the signal, where the latter property is facilitated

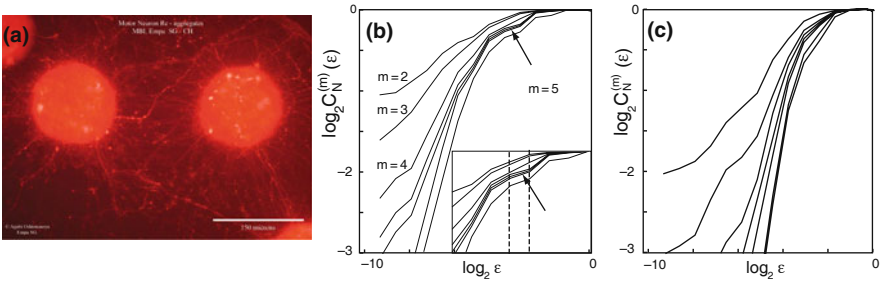


Fig. 8 (**a**) Motor neuron re-aggregates of spherical extension of 120–150 μm, cultivated on a 2-D micro-electrode array. (**b**) Signal with a temporal pattern, (**c**) essentially noise (see text)

[1] Data and Fig. 8a by courtesy of the Swiss Materials Science Institute EMPA of St. Gallen, Dr. A. Bruinink and A. Osh

by the log–log scale display. The proposed tool will be beneficial in particular in the biological, medical and applied technical sciences, where currently a strong focus is on the dissection of signals into noise and patterns and on multiscale modeling.

References

1. Z. Toroczkai, G. Károlyi, Á. Péntek, T. Tél, and C. Grebogi. *Phys. Rev. Lett*., **80**, 500 (1998).
2. S. Martignoli, J. J. van der Vyver, A. Kern, Y. Uwate, and R. Stoop. *Appl. Phys. Lett*., **91**, 064108 (2007).
3. R. Stoop and B. I. Arthur Jr. *Chaos*, **18**, 023123 (2008).
4. B. L. Strehler. *Perspect. Biol. Med*., **12**, 584 (1969).
5. J. E. Dayhoff, G. L. Gerstein. *J. Neurophysiol*., **49**(6), 1334 (1983).
6. J. E. Dayhoff and G. L. Gerstein. *J. Neurophysiol*., **49**(6), 1349 (1983).
7. R. Lestienne and B. L. Strehler. *Brain Res*., **437**, 214 (1987).
8. M. Abeles and G. L. Gerstein. *J. Neurophysiol*., **60**(3), 909 (1988).
9. R. Lestienne and H. C. Tuckwell. *Neuroscience*, **82**(2), 315 (1998).
10. Y. Prut, E. Vaadia, H. Bergman, I. Haalman, H. Slovin, and M. Abeles. *J. Neurophysiol*., **79**, 2857 (1998).
11. R. Stoop, K. Schindler, and L. A. Bunimovich. *Acta Biotheor*., **48**, 149 (2000).
12. I. V. Tetko and A. E. P. Villa. *J. Neurosci. Methods*, **105**, 1 (2001).
13. I. V. Tetko and A. E. P. Villa. *J. Neurosci. Methods*, **105**, 15 (2001).
14. R. Stoop, D. A. Blank, A. Kern, J.-J. van der Vyver, M. Christen, S. Lecchini, and C. Wagner. *Cog. Brain Res*., **13**, 293 (2002).
15. R. Stoop, K. Schindler, and L. Bunimovich. *Biol. Cybern*., **83**, 481 (2000).
16. M. Christen, A. Kern, A. Nikitchenko, W.-H. Steeb, and R. Stoop. *Phys. Rev. E*, **70**, 011901 (2004).
17. P. Grassberger and I. Procaccia. *Phys. D*, **13**, 34 (1984).
18. C. Grebogi, E. Ott, and J.A. Yorke. *Phys. Rev. Lett*., **37**, 1711 (1988).
19. J. Peinke, J. Parisi, O. E. Roessler, and R. Stoop. *Encounter with Chaos*. Springer, Berlin (1992).
20. H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge (2000).
21. A. Kern, W.-H. Steeb, and R. Stoop. *Z. Naturforsch*., **54**a, 404 (1999).
22. F. Takens. Detecting strange attractors in turbulence. In D. A. Rand and L. S. Young, editors, *Dynamical Systems and Turbulence*, Lecture Notes in Mathematics. Springer, Berlin, vol. 898, pp. 366–381 (1981).
23. D. L. Valladares, S. Boccaletti, C. Grebogi, and H. Mancini. *Int. J. Bifurcat Chaos*, **11**, 2621 (2001).
24. D. Yu, M. Small, R. G. Harrison, and C. Diks. *Phys. Rev. E*, **61**, 3750 (2000).

Index

M. Thiel et al. (eds.), *Nonlinear Dynamics and Chaos: Advances and Perspectives,*
Understanding Complex Systems, DOI 10.1007/978-3-642-04629-2,