Lecture Notes in Physics

The Editorial Policy for Monographs

The series Lecture Notes in Physics reports new developments in physical research and teaching - quickly, informally, and at a high level. The type of material considered for publication includes monographs presenting original research or new angles in a classical field. The timeliness of a manuscript is more important than its form, which may be preliminary or tentative. Manuscripts should be reasonably self-contained. They will often present not only results of the author(s) but also related work by other people and will provide sufficient motivation, examples, and applications.

Acceptance

The manuscripts or a detailed description thereof should be submitted either to one of the series editors or to the managing editor. The proposal is then carefully refereed. A final decision concerning publication can often only be made on the basis of the complete manuscript, but otherwise the editors will try to make a preliminary decision as definite as they can on the basis of the available information.

Contractual Aspects

Authors receive jointly 30 complimentary copies of their book. No royalty is paid on Lecture Notes in Physics volumes. But authors are entitled to purchase directly from Springer other books from Springer (excluding Hager and Landolt-Börnstein) at a $33\frac{1}{3}\%$ discount off the list price. Resale of such copies or of free copies is not permitted. Commitment to publish is made by a letter of interest rather than by signing a formal contract. Springer secures the copyright for each volume.

Manuscript Submission

Manuscripts should be no less than 100 and preferably no more than 400 pages in length. Final manuscripts should be in English. They should include a table of contents and an informative introduction accessible also to readers not particularly familiar with the topic treated. Authors are free to use the material in other publications. However, if extensive use is made elsewhere, the publisher should be informed. As a special service, we offer free of charge LaTeX macro packages to format the text according to Springer's quality requirements. We strongly recommend authors to make use of this offer, as the result will be a book of considerably improved technical quality. The books are hardbound, and quality paper appropriate to the needs of the author(s) is used. Publication time is about ten weeks. More than twenty years of experience guarantee authors the best possible service.

LNP Homepage (springerlink.com)

On the LNP homepage you will find:
−The LNP online archive. It contains the full texts (PDF) of all volumes published since 2000. Abstracts, table of contents and prefaces are accessible free of charge to everyone. Information about the availability of printed volumes can be obtained.
−The subscription information. The online archive is free of charge to all subscribers of the printed volumes.
−The editorial contacts, with respect to both scientific and technical matters.
−The author's / editor's instructions.

J.-R. Chazottes B. Fernandez

Dynamics of Coupled Map Lattices and of Related Spatially Extended Systems

Springer

Authors

Jean-René Chazottes
CNRS-Ecole Polytechnique
Centre de Physique Théorique
91128 Palaiseau CX, France
Email: jeanrene@cpht.polytechnique.fr

Bastien Fernandez
CNRS Luminy
Centre de Physique Théorique
FRUMAM B.P. 907
13288 Marseille CX 09, France
Email: bastien@cpt.univ-mrs.fr

J.-R. Chazottes, B. Fernandez, *Dynamics of Coupled Map Lattices and of Related Spatially Extended Systems*, Lect. Notes Phys. 671 (Springer, Berlin Heidelberg 2005), DOI 10.1007/b103930

Library of Congress Control Number: 2005923688

ISSN 0075-8450

ISBN-10 3-540-24289-9 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-24289-5 Springer Berlin Heidelberg New York

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

Printed in The Netherlands

Typesetting: by the authors and TechBooks using a Springer LaTeX macro package

Printed on acid-free paper SPIN: 11360810 57/TechBooks 5 4 3 2 1 0

Preface

This book assembles the notes that have been written for the lectures delivered at the school-forum CML2004, a meeting held at Institut Henri Poincaré (Paris) from the 21st of June to the 2nd of July 2004.

The purposes of CML2004 were to give a survey of the most significant achievements in the theory of coupled map lattices (CML) and of related spatially extended systems, and to stimulate debates on open problems and future directions of research.

The panorama of CML theory has been presented through a series of 15 lectures given by specialists. Additional results have been given in the form of short communications, of posters and of young scientist sessions (organised by Pierre Guiraud and Arnaud Meyroneinc). In addition to these expository sessions, stimulated discussions and public debates on questions and on relevant open problems have taken place.

About 60 participants attended the meeting. They essentially consisted of physicists and mathematicians in nonlinear dynamics, from young scientists (about 20 participants where PhD students or postdocs) to senior scientists, working either on CML or on other subjects (about 20 participants).

As *Ecole Thématique*, CML2004 received essential support from the CNRS without which the meeting could not have been realised. The following institutions also contributed to the financial support: *Ministère de la Recherche et des Nouvelles Technologies*, *Centre de Physique Théorique de Palaiseau*, *Centre de Physique Théorique de Marseille*, *École Polytechnique*, *Ministère des Affaires Etrangères*, *Université de Provence*, *École doctorale de Marseille "Physique et Sciences de la Matière"*, *European Physical Society (Young Physicist Fund).*

The participants appreciated the facilities of the Institut Henri Poincaré. The organisers appreciated the assistance of its administrative staff. Arkady Pikovsky, who could not attend CML2004, helped us to raise funds. We are grateful to him for his advice. As a special acknowledgement, we wish to express our gratitude to Ricardo Lima who has kept encouraging and supporting us since the beginning of our respective post-graduate studies.

Paris, Marseille — *Jean-René Chazottes*
November 2004 — *Bastien Fernandez*

List of Contributors

Valentin Afraimovich[1]
IICO–UASLP
Alvaro Obregón 64, San Luis Potosí
SLP 78000, México
valentin@cactus.iico.uaslp.mx

Peter Ashwin[1]
Dept. of Mathematical Sciences
Laver Building, University of Exeter
Exeter EX4 4QE, UK
P.Ashwin@ex.ac.uk

Claude Baesens[1]
Mathematics Institute
University of Warwick
Coventry CV4 7AL, UK
claude@maths.warwick.ac.uk

Leonid A. Bunimovich[1]
Georgia Institute of Technology
School of Mathematics
Atlanta, GA 30332, USA
bunimovh@math.gatech.edu

Jean-René Chazottes
Centre de Physique Théorique
CNRS & École Polytechnique
91128 Palaiseau Cedex, France
jeanrene@cpht.polytechnique.fr

Ricardo Coutinho[1]
Departamento de Matemática
Instituto Superior Técnico
Av. Rovisco Pais 1096, Lisboa Codex
Portugal
Ricardo.Coutino@math.ist.utl.pt

Bard Ermentrout[1]
Department of Mathematics
University of Pittsburgh
Pittsburgh, PA 15260, USA
bard@euler.math.pitt.edu

Bastien Fernandez
Centre de Physique Théorique
CNRS, Luminy, Case 907
13288 Marseille Cedex 09, France
bastien@cpt.univ-mrs.fr

Luis Mario Floría[1]
ICMA
Dpt. of Theory and Simulation of
Complex Systems
CSIC-Universidad de Zaragoza
E-50009 Zaragoza, Spain
floria@posta.unizar.es

Jesus Gómez-Gardeñes
ICMA
Dpt. of Theory and Simulation of Complex Systems
CSIC-Universidad de Zaragoza
E-50009 Zaragoza, Spain
gardenes@unizar.es

Hidde de Jong[1]
INRIA
Unité de recherche Rhône-Alpes
655 avenue de l'Europe, Montbonnot
38334 Saint Ismier Cedex, France
Hidde.de-Jong@inrialpes.fr

Esa Järvenpää[1]
Dept. of Mathematics and Statistics
P.O. Box 35 (MaD)
FIN-40014 University of Jyväskylä
Finland
esaj@maths.jyu.fi

Wolfram Just[1]
MRC
Queen Mary / University of London
London, UK
w.just@qmul.ac.uk

Gerhard Keller[1]
Mathematisches Institut
Universität Erlangen-Nürnberg
Bismarckstr. 1 1/2
91054 Erlangen, Germany
keller@mi.uni-erlangen.de

Ricardo Lima[1]
Centre de Physique Théorique
CNRS, Luminy, case 907
13288 Marseille Cedex 09, France
lima@cpt.univ-mrs.fr

Carlangelo Liverani
Dipartimento di Matematica
Università di Roma II (Tor Vergata)
Via della Ricerca Scientifica
00133 Roma, Italy
liverani@mat.uniroma2.it

Robert S. MacKay[1]
Mathematics Institute
University of Warwick
Coventry CV4 7AL, UK
mackay@maths.warwick.ac.uk

Yuri L. Maistrenko[1]
Institute of Mathematics
Academy of Sciences of Ukraine
01601 Kyiv, Ukraine
maistrenko@imath.kiev.ua

Oleksandr V. Popovych
Institute of Mathematics
Academy of Sciences of Ukraine
01601 Kyiv, Ukraine
o.popovych@fz-juelich.de

Frank Schmüser
IAKS, University of Karlsruhe
Karlsruhe, Germany
schmues@ira.uka.de

Peter A. Tass
Institute of Medicine
Research Centre Jülich
52425 Jülich, Germany
p.tass@fz-juelich.de

[1] Lecturer at CML2004

Contents

The CML2004 Project

J.-R. Chazottes[1] and B. Fernandez[2]

[1] Centre de Physique Théorique, CNRS & École Polytechnique 91128 Palaiseau Cedex, France
jeanrene@cpht.polytechnique.fr
[2] Centre de Physique Théorique, CNRS & Universités de Marseille et Toulon Luminy Case 907, 13288 Marseille Cedex 09, France
bastien@cpt.univ-mrs.fr

Coupled map lattices (CML) are basic models for the time evolution of nonlinear systems which, above all, are extended in space or involve many individual units. The characteristic features of CML are

- discrete time dynamics
- discrete nature of the underlying space (lattice or network)
- the local variables consist of real numbers or real vectors.

Formally speaking, a CML is a discrete time dynamical system generated by a mapping acting on real (vector) sequences. The typical and most studied example is the model introduced by Kaneko in 1983 and given by the following iterations

$$u_s^{t+1} = (1-\epsilon)f(u_s^t) + \frac{\epsilon}{2}(f(u_{s-1}^t) + f(u_{s+1}^t)) \quad t \in \mathbb{N},\ \epsilon \in [0,1]$$

where $u_s^t \in \mathbb{R}$ and f is a real mapping.

Depending on the context, the configurations $\{u_s^t\}$ represent the spatial profile of a chemical concentration, of a population density, of a velocity field, etc. In these cases, the configurations are bounded sequences, sometimes finite or periodic. Some systems however require unbounded configurations. This is the case for instance in the Frenkel-Kontorova model of particle chains where u_s^t represents the position along the real line of the sth particle, see the chapters on monotone dynamics [Floría, Baesens and Gómez-Gardenes], [Baesens] and [Coutinho].[1]

As shown by the basic model, the dynamics of a CML is governed by two competing terms; an individual nonlinear reaction represented by f and a spatial interaction (coupling) with variable intensity ϵ. In the basic model, the interaction is a convolution operator which represents a diffusive coupling. These two terms are applied successively, a characteristic feature of CML, and

[1] We refer to the chapters by using the name(s) of their author(s).

J.-R. Chazottes and B. Fernandez: *The CML2004 Project*, Lect. Notes Phys. **671**, 1–8 (2005)
www.springerlink.com

this ensures that bounded initial configurations remain bounded (provided that f is bounded). However, CML are not restricted to such composition nor to convolution couplings and many other models have been considered.

Their simple formulation make CML a paradigm of nonlinear spatially extended dynamical systems. In particular, CML are specially designed to facilitate computer simulations over large space-time domains. The simulations exhibit an extraordinary large panel of behaviours upon changes in the local map and in the interaction (or simply in their parameters). This diverse phenomenology motivated the application of CML to the simulation of real systems. For instance, a spectacular application pointed out to us by Pierre Guiraud is afforded by the simulation of cloud formation by a CML derived from fluid dynamics equations[2].

In the endeavour to describe CML analytically, various methods, techniques and tools have been borrowed from the theory of Dynamical Systems (stability analysis, Lyapunov exponents, bifurcations, symbolic dynamics, etc.). Some results have been confirmed or obtained in a rigorous mathematical framework (e.g. global and partial synchronisation, front dynamics, etc.). As far as Mathematics is concerned, CML form a proper source of problems since they are dynamical systems with infinite dimensional phase space and since they do not satisfy the usual assumptions on dynamics for large physically relevant sets of parameters (e.g. uniform hyperbolicity, prescribed symbolic dynamics, etc.).

For a more complete exposition of the origins of CML, of their motivations, and for an overview of problems, we refer to chapter introductions, especially of [Bunimovich] and [Just and Schmüser].

The purposes of the meeting CML2004 were to present a survey of the theory of CML and of related spatially extended systems (lattice dynamical systems, discrete time systems with continuous space, integro-differential equations, etc.), and to stimulate debates on open problems and future directions of research. In order to cover both physical and mathematical aspects, to avoid overlap between lectures and to appeal to a broad audience, 15 specialists were invited to present results on a given theme. By doing so, we were conscious of the fact that many significant contributors to the theory of CML could not present their results. But we had the feeling that a limited number of lectures could bring more material to a large audience than a series of talks.

This volume collects the notes written by the lecturers, sometimes with the help of collaborators. The themes cover numerical, theoretical and mathematical aspects of various spatially extended systems. More than the results themselves, concepts, techniques and tools developed for their analysis are presented. Since the investigation of a model on its own without any relationship to concrete situations has only little interest, examples of comparison and of

[2] Go to `http://nis-lab.is.s.u-tokyo.ac.jp/~nis/animation.html` to see the movie *Cloud simulation by CML* and to download the related paper [R. Miyazaki, S. Yoshida, Y. Dobashi and T. Nishita, Proc. of Pacific Graphics (2001) 363–372)].

adaptation to physical and biological problems are given. The presentation is by no means complete, but we hope it can serve as a basis for future research on spatially extended systems.

Before going into details we present a schematic overview of most significant phenomena in Fig. 1. This picture collects the dynamical regimes, together with the transitions between regimes, which occur depending on the local map and on the interaction intensity. Naturally, the phenomenology here has been fairly simplified. It does not make any distinction between various forms of interaction (local coupling or global coupling). More importantly, it does not make any distinction between various spatially extended systems. To a smaller extent, neither does it take into account the lattice size dependence nor the dependence on boundary conditions.[3] Still we hope that this figure can guide the reader through the book.

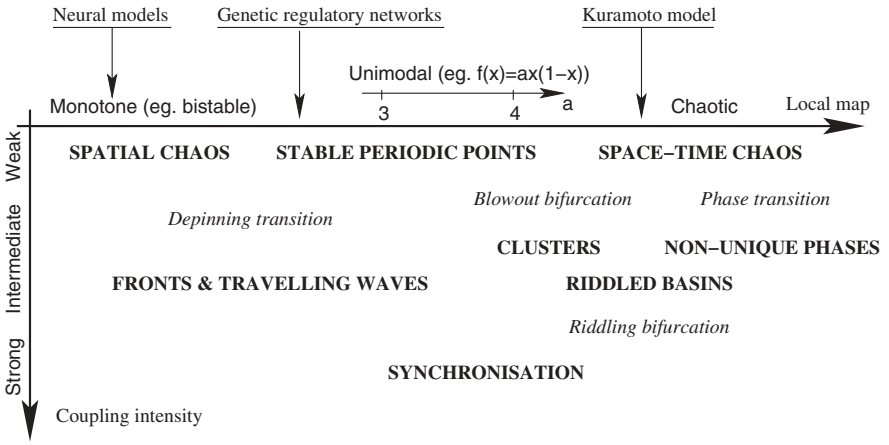


Fig. 1. Schematic representation of the phenomenology of CML (and of related models) *versus* the local map and the coupling intensity. Obviously, the phenomena may extend to larger domains than those indicated here. For instance, synchronisation may occur for any local map provided that the coupling (and the lattice size) is suitably chosen

The chapters have been assembled into 4 thematic parts. The first two parts are devoted to the description of statistical and geometric properties of CML. The third part collects results on the dynamics of monotone spatially extended systems. The last part concerns the introduction and analysis of models motivated by dynamical problems in Biology.

[3] Apart from an example in Sect. 5 of [Bunimovich], the effect of boundary conditions remain largely unravelled.

1 Statistical Properties of Coupled Chaotic Maps

Inspired by the analogy with spin models in Statistical Mechanics which emerges from symbolic dynamics, the consideration of global statistical properties of chaotic CML started soon afterward the introduction of CML, in 1988 precisely. The analogy suggests that, typically, phase transitions should occur when the coupling parameter increases. The transition is expected to split a unique space-time chaotic phase (high temperature) into several ordered phases (low temperature). However, the reputation of Statistical Mechanics technical difficulties warns that any attempt on a rigorous description of a phase transition in CML would face arduous problems.

To start with, characterising space-time chaotic phases is a problem in its own which has been the preliminary focus of many studies. A mixing hyperbolic dynamical system on a compact set has a (unique) natural phase with several equivalent characteristic properties. This equivalence fails in infinite lattices. Characterising the natural measures then needs to be addressed prior to any other statistical property in CML.

Various proposals have been made. Using again analogy with Statistical Mechanics, a natural measure should be the Gibbs measure of an appropriate Hamiltonian on the space-time lattice [Just and Schümser], [MacKay] and [Jarvenpää]. In the framework of the theory of Dynamical Systems and with an ergodic theorem in mind, a natural measure should describe the statistics of orbits issued from "typical" initial conditions, [Bunimovich] and [Keller and Liverani]. In the dual formulation of the dynamics, a natural measure should be the limit of iterations of any "regular" initial distribution, [Bunimovich], [MacKay] and [Keller and Liverani].

With a definition provided, the question of uniqueness of the natural measure in infinite lattices comes to the centre of attention. Contrary to the case of finite lattices of weakly coupled chaotic maps, requiring that all finite dimensional projections be absolutely continuous does not ensure uniqueness [Jarvenpää]. On another hand, due to infinite extension, some transients may last forever and can thus be defined as a proper phase; a phenomenon which does not exist in finite lattices [Just and Schmüser] and [MacKay]. Uniqueness can be shown however for small couplings in a suitable Banach space (of measures having finite dimensional marginals with at most exponentially growing total variation) [Keller and Liverani].

Two distinct approaches to phase transitions have been proposed. One approach is based on the formal derivation of a master equation for probabilities associated with atoms of the symbolic partition. It consists in showing that some transitions between atoms depend on the coupling parameter [Just and Schmüser]. However, this approach can be hardly controlled from a mathematical point of view. More importantly, changes in transition probabilities correspond to changes in the topology of the repeller (bifurcation) rather than to changes in its statistics only. Such changes may not be due to infinite spatial extension but may also occur in finite lattices. In this case, the term "phase

transition" would not be appropriate [Bunimovich], [Just and Schmüser and [MacKay].

An alternative mathematically rigorous approach is to construct CML with prescribed phase transitions. The CML consist in piecewise affine mappings based on probabilistic cellular automata (PCA) which have been proved to possess a phase transition, in particular Toom's PCA [Just and Schmüser] and [MacKay]. Coupling there is introduced by letting the local map depend on symbolic states at neighbouring sites. Such models fairly differ from classical CML. However, this trick allows to overcome the unsolved problem of determining the symbolic dynamics of a CML for an arbitrary coupling parameter.

2 Geometric Aspects of Lattice Dynamical Systems

Beside focusing on specific phenomena such as phase transitions, and as suggested by the explicit dependence on the coupling parameter, a standard issue in CML is to describe the dynamics over the entire coupling parameter range. Due to competitions between local and interaction terms and between linear and nonlinear terms, this is a formidable task which has been accomplished only in particular cases. In arbitrary lattices, only the extreme regimes of weak and of strong couplings can be considered as satisfactorily described.

In view of perturbation theory, the dynamics at each site in a CML with weak coupling can be regarded as a local map perturbed by contributions from other sites. Accordingly, the behaviours in uncoupled and in weakly coupled regimes should be qualitatively the same provided that the local map dynamics is robust to perturbations.

The simplest case is when the local map has two stable fixed points. Then if the coupling parameter is small enough, just as in the uncoupled case the CML has an infinite set of stable fixed points on which the action of space translations has positive topological entropy [Afraimovich]. This property is called spatial chaos and extends to weakly coupled lattices of local maps with stable periodic orbits.

When the local map is strongly chaotic, space-time chaos exists for small coupling. That is to say, when the local map has a hyperbolic set with positive topological entropy, then the CML with sufficiently weak coupling has a hyperbolic set on which the $\mathbb{Z}^2$-action of space-time translations also has positive topological entropy.

In spite of being intuitively simple, weak interaction regimes gave the opportunity to adapt to lattice systems various dynamical systems techniques, e.g. persistence of uniform hyperbolicity under weak coupling, symbolic dynamics [Afraimovich] and [MacKay]. They also allowed to obtain results which are specific to lattice dynamical systems, e.g. description of space-time periodic configurations as orbits of a low-dimensional map, density of space-time (quasi-)periodic configurations with given (quasi-)period [Afraimovich].

Whereas structural stability of uncoupled systems does not depend on the symmetry of translation invariance and extends to some heterogeneous CML, in strongly coupled regimes, the dynamics relies on this symmetry.

The basic strongly coupled regime in translation invariant CML is synchronisation. In this context synchronisation means that the subset of constant configurations, namely the diagonal, attracts all orbits in phase space [Afraimovich] and [Maistrenko, Popovych and Tass].

Synchronisation takes place when all transverse eigenvalues of the mapping derivative computed at any point on the diagonal have modulus uniformly smaller than 1. The synchronisation is said to be chaotic if the tangential Lyapunov exponent on the diagonal (which in CML is nothing else but the Lyapunov exponent of the local map) is positive.

In the case where only the transverse Lyapunov exponents on the diagonal are negative (which happens when the coupling parameter decreases from the synchronisation regime) the basin of attraction of the diagonal is only local and may have a complex riddled structure, a phenomenon called partial synchronisation. Riddled basins are not limited to CML but emerge in a broader context, specifically in equivariant dynamical systems [Ashwin].

Riddled basins only concern a neighbourhood of the diagonal. The rest of phase space may contain orbits not asymptotically approaching the diagonal, a reminiscence of weakly coupled regimes. A simple example is a stable periodic orbit [Maistrenko, Popovych and Tass]. An example with a dense subset of unstable periodic orbits has also been exhibited [P. Glendinning, *Milnor attractors and topological attractors of a piecewise linear map*, Nonlinearity **14** (2001) 239–257].

3 Spatially Extended Systems with Monotone Dynamics

The typical situation for which the dynamics of a spatially extended system can be reasonably analysed over the whole coupling parameter range is that of systems with monotone dynamics. If an initial configuration lies below another initial configuration, then this ordering is preserved at later times.

Monotonicity is a classical property in parabolic partial differential equations (maximum principle). In lattices of coupled ordinary differential equations, it follows from cooperativity [Baesens]. For instance, it holds in the paradigmatic Frenkel-Kontorova model when assuming strong enough dissipation. In CML monotonicity holds for every $\epsilon \in [0, 1]$ provided that the local map f is an increasing function [Coutinho and Fernandez].

With the dynamics of chains in periodic potentials in mind, monotonicity can be completed with translation invariance and periodicity. Periodicity means that if the difference between two initial configurations equals, say 1 at all sites, then this difference remains unchanged at later times.

A monotone periodic translation invariant system has a regular and uniform nonlinear dynamics. Either each orbit remains sandwiched between stationary configurations (pinned regime) or all orbits indefinitely increase (or decrease) with finite velocity (sliding regime). In the sliding regime the propagation velocity is unique in phase space and there are corresponding travelling waves with rotationally ordered shape. The propagation velocity continuously depends on the system parameters. In particular, it is known to be positive for sufficiently large driving force (sufficiently large local map asymmetry in CML).

This phenomenology does not depend on the details of the model, a system either with continuous time [Baesens] or with discrete time. Neither it does depend on the details of the spatial interaction (discrete or continuous diffusion or both) [Coutinho and Fernandez]. This justifies substituting certain models by more convenient ones. In particular, one may assume the dynamics of lattice systems with small step sizes (discrete diffusion) to be suitably represented by the dynamics of a system defined on the whole real line (continuous diffusion), or vice-versa.

Excepted when generated by a driving force, transport may also be caused by a time dependent action on the system (non-autonomous system). In particular, switching on and off an asymmetric potential or switching on and off the diffusive interaction may also generate propagation (ratchet effect) [Floría, Baesens and Gómez-Gardeñes].

Even though the orbits remain bounded between two stationary configurations, there may be propagation. In this case, propagation concerns interfaces (discommensurations) between two contiguous stable stationary configurations. (Interfaces between a stable and an unstable configuration can also be relevant.) Bistable systems provide the basic framework where propagation of interfaces between stable phases can be analysed [Coutinho and Fernandez].

Bistable spatially extended systems satisfy monotonicity, translation invariance and the existence of two stable constant configurations at distance, say 1. In discrete time systems with arbitrary spatial interaction, the dynamics of interfaces is analogous to the previous one. There exists a unique asymptotic horizontal velocity for all interface orbits, this velocity depends continuously on the system parameters, and there are travelling waves (fronts).

4 Specific Lattice Dynamical Systems

In certain lattices with few sites, the dynamics can be described in the whole coupling range even though monotonicity is not assumed. A typical example with rich phenomenology is the Kuramoto model of globally coupled oscillators, a system of coupled ordinary differential equations. In this model, the sequence of bifurcations generated by decreasing the coupling is well-established [Maistrenko, Popovych and Tass]. Starting from a globally attracting synchronised orbit, bifurcations split asymptotic configurations into

clusters with synchronised motions. For smaller couplings, clusters break into independent oscillators. In lattices with a large enough number of sites, this scenario includes a chaotic attractor for intermediate couplings. Motivated by synchronisation caused neurological diseases in brain function, the Kuramoto model has been employed to simulate the impact of a stimulation on an assembly of neurons.

The mechanisms leading to synchronisation in networks of neurons have been thoroughly investigated taking into account detailed neurons and synapses characteristics [Ermentrout]. In a different context where the neurons have a excitable dynamics and not an oscillatory one, propagating waves with regular or lurching motion have been exhibited. Some of these waves are the analogous of travelling fronts in systems with monotone dynamics mentioned above.

Another class of biological systems which comprehension involves network dynamics is that of genetic regulatory networks. The dynamical characteristics of the mechanisms involved in this context are a local piecewise contracting dynamics combined with a complex interaction graph [de Jong and Lima]. This combination is rather original in the framework of lattice dynamical systems and the resulting dynamics has only been completely described in networks with simple graphs.

Coupled Map Lattices: at the Age of Maturity

L.A. Bunimovich

Georgia Institute of Technology, School of Mathematics, Atlanta, GA 30332 USA
bunimovh@math.gatech.edu

1 Introduction

Coupled Map Lattices (CML) were simultaneously and independently introduced by K. Kaneko, R. Kapral and S. Kuznetsov in 1983–84 [1, 2, 3, 4, 5, 6]. CML describe the time evolution of fields that can be split into an independent evolution of local systems (elements) of these fields (usually defined by some map of a local phase space) followed by (spatial) interactions of these local systems generated by some operator acting on the entire (global) phase space of CML. A structure of local systems in CML forms a lattice. At any moment of time all values of local variables are defined. These values determine a spatial structure (pattern) of the field. In CML it is assumed that all local dynamical systems are identical and that the spatial interactions between any local system and the rest of CML are the same for all local systems. (In more general Lattice Dynamical Systems (LDS) dynamics is not assumed to be a composition of a local dynamics and spatial interactions and neither to be translationally invariant.)

It is a remarkable fact that three scientists in three different countries introduced the same model, while having quite different goals. Kapral used CML to model the dynamics of spatially extended systems in chemistry, Kuznetsov was interested in developing a renormalization group approach similar to Feigenbaum's universality to spatially extended systems (especially to electrical circuits). Kaneko had quite different and more general agenda. He considered CML to be a new powerful model to study dynamics of spatially extended systems of virtually any nature and was the most active researcher and propagandist in this area (see e.g., [7, 8, 9, 10] and references therein).

It is believed that there were two major reasons for the introduction of CML. In the early 80s the theory of finite dimensional (point, non-extended) dynamical systems had been already fairly well developed during the preceding two decades started with the seminal Kolmogorov's paper [11] which built

L.A. Bunimovich: *Coupled Map Lattices: at the Age of Maturity*, Lect. Notes Phys. **671**, 9–32 (2005)
www.springerlink.com

a bridge between deterministic and probabilistic dynamics. However, just a very moderate success had been achieved in the studies of the dynamics of spatially extended systems, traditionally described by partial differential equations (PDE). First, CML provided seemingly simpler models where one can usually safely assume that dynamics of local components of the field (local dynamics systems or local maps) is (well) known and therefore the effect(s) of spatial interactions could be singled out. And secondly, CML are, in a sense, ideal models for computer simulations. Indeed, CML are discrete in space. Therefore, they allow a direct numerical simulation without using various discretization schemes required by PDE. This crucial feature of CML was especially intensively used by Kaneko, who produced a large variety of numerical studies of CML and provided a lot of pictures which visualized spatio-temporal patterns of different regimes of dynamics [7, 12]. At this initial stage CML started to attract a lot of attention not just as useful computational models but also as promising phenomenological models of space-time dynamics.

Mathematical studies of CML started with the paper [13], which developed an approach to a rigorous investigation of these systems. This approach is based on an extension of the thermodynamic formalism developed by Sinai, Ruelle and Bowen [14, 15, 16] for finite dimensional (local) dynamics to spatially extended systems, especially to CML.

The thermodynamic formalism is a far reaching extension of symbolic dynamics. Symbolic dynamics is based on a coding of orbits of dynamical systems into double-infinite (in case of invertible dynamics) or semi-infinite (in case of non-invertible dynamics) sequences of positive integers. By virtue of such coding, the dynamics gets represented in the corresponding space of (admissible by dynamics) sequences by their shift on one coordinate. The thermodynamic formalism goes much farther by interpreting these sequences as configurations of spins (discrete variables) of particles in lattice systems of statistical mechanics. The corresponding potential of interaction between these (virtual) particles (or spins) is determined by the (initial) dynamical system and equals logarithm of its Jacobian (see e.g., [15]).

Statistical mechanics consists of two parts, that are called equilibrium and nonequilibrium statistical mechanics, respectively. The equilibrium statistical mechanics deals with stationary and equilibrium measures for the corresponding systems. The central problem in equilibrium statistical mechanics is to find all equilibrium measures. These equilibrium measures are called phases. If a number of equilibrium measures changes (with a change of some parameter, usually a temperature) then the system is said to have a phase transition. The corresponding value of the parameter, where a phase transition occurs is called a critical value (e.g., a critical temperature).

In the paper [13] (see also [17]) it has been conjectured that if the local systems in CML are strongly chaotic and if the spatial interactions are sufficiently weak, then in the corresponding system there is only one equilibrium measure. Therefore this conjecture claims that there are no phase transitions in the range of weak spatial interactions provided that local systems are strongly

chaotic (e.g., uniformly hyperbolic). This conjecture has been proved in [13] for a class of expanding one-dimensional (local) maps with a special type of spatial interactions which provides a Markov partition for such CML. This result has been then proved for much more general local dynamics and less restrictive classes of spatial interactions [19, 31].

It has also been shown in [13] (as well in [19, 31]) that the corresponding CML with this unique equilibrium measure has the property of space-time mixing, i.e., the correlations in these systems decay both in time and in space. This property of space-time mixing is one of the strongest properties of space-time chaos that a spatially extended system may have. However, it is just one out of many various properties of space-time chaos in spatially extended dynamical systems. Unfortunately this property of space-time mixing has been called in [13] a space-time chaos, and despite the attempts to correct this terminology [32] it is still often addressed as space-time chaos.

The absence of phase transitions and space-time mixing mean that all typical (with respect to this unique equilibrium measure) orbits of the corresponding CML behave and look similar, i.e., they produce essentially the same spatio-temporal patterns. This property of absence of different spatio-temporal patterns in an extended system has been interpreted as the *absence of coherent structures* in such systems [13, 17].

The presence of coherent structures is a well known phenomenon in extended dynamical systems. It has been first observed and received this name in experimental studies of turbulence with high Reynolds numbers. Before the paper [13] there was no general definition of coherent structures. Instead, this name was usually given to some simple spatial or spatio-temporal patterns that have been observed in real or numerical experiments or to some simple exact solutions of dynamical equations of spatially extended systems.

In [13] not only a general definition of coherent structures has been given but also a principally new interpretation of this phenomenon has been presented. This interpretation is based on the correspondence between spatially extended dynamical systems and lattice models of statistical mechanics provided by the thermodynamic formalism.

A coherent structure according to [13] is an invariant measure μ of a spatially extended system. A support of an invariant measure is formed by some set of orbits of a spatially extended dynamical system. If this measure is ergodic then all these orbits have the same spatio-temporal structure. This set of similar in this sense orbits forms a coherent structure corresponding to the measure μ. (Usually, a dynamical systems has many invariant measures but only a few of them, that are in some sense "observable" (realizable) in experiments of various nature are of interest. These measures are referred to as physical, stable, natural, etc. A choice of relevant measure(s) is crucial in dealing with statistical properties of dynamical systems, and we will discuss it later in details.)

A measure having the property of space-time mixing is, of course, ergodic. If an extended dynamical system has a unique (observable) measure with the space-time chaos then, just by definition of space-time chaos, its (typical) orbits have no coherence whatsoever. However, if an extended system has several equilibrium (observable) measures then the orbits in the supports of such measures already have some coherence. Indeed, a choice of one of these measures provides some information about properties (structure) of the orbits (solutions) that belong to the support of this measure.

Therefore, the approach developed in [13] allows to relate coherent structures of spatially extended dynamical systems to phase transitions in the corresponding to these systems models of statistical mechanics. After phase transitions several equilibrium measures appear and orbits in their support are not space-time chaotic. Hence, they have some kind of coherence. Observe, that such orbits still could be rather chaotic, but they are not "completely" chaotic. Thus, the treatment of coherent structures as measures is rather general. It allows, for instance, to compare a level of coherence of different structures (measures). Many efforts were made to verify this approach to coherent structures numerically and to justify it rigorously [33, 48]. This area remains one of the most exciting in the studies of lattice dynamical systems.

Phase transitions occur when some parameter(s) of spatially extended dynamical systems change. Therefore phase transitions in this sense are somewhat similar to bifurcations of dynamical systems. This similarity could be helpful, but, on another hand, it already led to a lot of confusions. Especially, it became kind of fashionable to call bifurcations phase transitions. (I believe that one of the reasons for that is a sort of feeling that phase transitions are better funded now than bifurcations. Although this feeling may be true this confusion does not help in studies of this complicated and still not well defined and by far not well understood area.)

Certainly the major difference between spatially extended dynamical systems and nonextended (point) ones is in the presence of spatial interactions (between local (point) dynamical systems). Therefore (one of) the most important parameters is a strength (amplitude) of spatial interactions. Most of mathematical results on CML were obtained for weak spatial interactions. Then the strength of spatial interactions can be treated as a small parameter, and a problem can be analyzed by a kind of perturbation theory. Often this perturbation problem is highly nontrivial (e.g., singular) though.

There are three major areas where this approach has been applied. The first, already mentioned, deals with CML with strongly chaotic (e.g., mixing) local dynamics. Then such system obviously is space-time mixing when there are no spatial interactions at all between the local systems. Indeed, in such system the spatial correlations are identically zero. The second area deals with the construction of exact interesting solutions (often of some complicated nature) for corresponding infinite dimensional dynamical systems [49, 50, 51]. Usually the existence of these solutions is already known for CML without spatial interactions. To find such solutions one just needs to know that the

corresponding projections on the local dynamics exist. The problem is to prove a version of the implicit function theorem in an infinite dimensional space with properly defined norm. This area produced a variety of interesting results including the existence of breathers in some systems appeared in physics and other applications. The third area deals with the problem of construction and stability of more or less standard (e.g., waves) or chaotic solutions in LDS and/or with their hyperbolicity. A mathematical problem here again is in infinite dimension and therefore in a proper choice of a norm or metric [52, 53, 54, 55].

It is important to mention that the goal of the first two areas is to demonstrate that (at least in some cases) the dynamics of weakly interacting CML or LDS is qualitatively the same as of noninteracting systems. In other words these areas deal with situations where spatial interactions, that are at the heart of dynamics of spatially extended systems and, in fact, make these systems to be spatially extended, are in a sense negligible. The third area deals mainly with problems that are standard for nonextended (point) systems as well.

To understand the dynamics of spatially extended systems one should rather address a question how spatial interactions can influence (change) dynamics of local systems. One of such crucial problems is the already discussed problem of phase transitions, where a number of equilibrium states (measures) changes at some critical value of an amplitude of spatial interactions. Another question of this type deals with the analysis of bifurcations that occur when this amplitude changes, i.e., the amplitude of spatial bifurcations is considered as a bifurcation parameter.

The last problem is a central one in the analysis of interacting dynamical systems of any kind, i.e., for systems with finite as well as with infinite number of local components, for systems with identical local components (e.g., for CML) as well as for systems with different local components (e.g., for LDS of general type). It is worthwhile to mention that the classes of spatially extended or of interacting systems are naturally restricted when one addresses a specific problem. For instance, some authors insist that the notion of phase transitions should be used only for infinite systems (i.e., for systems with infinitely many interacting components) while the problem of synchronization usually deals only with systems with a few interacting components (subsystems) which are nonidentical. In fact, the phenomenon of synchronization assumes that (weak) interactions force different systems to evolve in some sense synchronously [56].

2 Multicomponent Dynamical Systems (MDS)

In this section we give the definitions of some types of dynamical systems with interacting components. Let $\mathcal{N}$ be a finite or countable collection of indices. Consider a direct product $(X^{\mathcal{N}}, \rho^{\mathcal{N}})$ of compact metric spaces (X_i, ρ_i),

$i \in \mathcal{N}$. Then $X^{\mathcal{N}} := \bigotimes_{i\in\mathcal{N}} X_i$ is a phase space of a multicomponent dynamical system, and X_i, $i \in \mathcal{N}$, are phase spaces of local components (which are sometimes called local phase spaces). Points $\bar{x} \in \mathcal{N}$ can be naturally represented as (finite or infinite) collections $\bar{x} = (x_i)$, $i \in \mathcal{N}$. A metric in $\mathcal{N}$ can be defined in different ways. The most popular (and in a sense natural one) is $\rho^{\mathcal{N}}(\bar{x}, \bar{y}) := \max_{i\in\mathcal{N}} \rho_i(x_i, y_i)$.

The collection T_i, $i \in \mathcal{N}$, of local maps $T_i : X_i \to X_i$ defines the local dynamics of a multicomponent system. We assume that any local system (component) is equipped with a reference measure m_i and that all local maps T_i are nonsingular with respect to m_i, $i \in \mathcal{N}$.

Often m_i is assumed to be the Lebesgue measure on X_i, $i \in \mathcal{N}$. It is a natural assumption for numerical (as well as for interpretation of real) experiments. Then a natural class of (initial, not steady state) probability distributions on X_i is formed by measures on X_i that are absolutely continuous with respect to m_i.

The local dynamics T_i defines the evolution of (probabilistic) measures μ on X_i via the relation

$$T_i^* \mu(A) = \mu(T_i^{-1} A) \tag{1}$$

where $A \subset X_i$ is a (measurable) subset of X_i. Therefore we can assume that T_i acts on (X_i, ρ_i, m_i), $i \in \mathcal{N}$.

Let $m^{\mathcal{N}} := \bigotimes_{i\in\mathcal{N}} m_i$. Consider a (nonsingular) map $T^{(\mathcal{N})} : X^{\mathcal{N}} \to X^{\mathcal{N}}$.

Definition 2.1. The pair $(X^{\mathcal{N}}, \mathcal{T}^{(\mathcal{N})})$ is called a *multicomponent dynamical system* if the dynamics $T^{(\mathcal{N})}$ can be decomposed into a superposition (composition) $\mathcal{T}^{(\mathcal{N})} = \mathcal{I}^{(\mathcal{N})} \circ T^{\mathcal{N}}$ of an interaction $\mathcal{I}^{(\mathcal{N})}$ between the local components and the direct product $T^{\mathcal{N}} := (\bigotimes_{i\in\mathcal{N}} T_i)$ of local maps T_i, $i \in \mathcal{N}$.

So far we did not assume that the collection of local components form any kind of (global) structure, i.e., we considered the indices $i \in \mathcal{N}$ as just abstract numbers. If, however, we consider this lattice of integers to be placed into some ("physical") space then the local components (and the corresponding variables x_i) form a lattice (field) in such space. Then we can consider the collection of values (x_i), $i \in \mathcal{N}$, as forming some spatial structure (snapshot) in the corresponding physical space. Of course, the set of indices $i \in \mathcal{N}$ does not need to be a (sub)set of integers. The indices could be integer vectors, and then we get a multidimensional lattice, elements of some set of a Cantor type, etc. In many cases, e.g., in the most of applications, the number of local components is finite.

One of the most amenable for theoretical studies and quite interesting for many applications (as well as in the situation when there is just a finite number of local components) is the (spatially) homogeneous case when all local systems (X_i, ρ_i, m_i, T_i), $i \in \mathcal{N}$, coincide.

We assume that the map $\mathcal{I}^{(\mathcal{N})} : X^{\mathcal{N}} \to X^{\mathcal{N}}$, which defines 'interactions' between local components is identical on the diagonal set $X^{\mathcal{N}}_{\mathrm{diag}} = \{\bar{x} \in X^{\mathcal{N}} : x_i = x_j \ \forall\, i, j\}$. This assumption means that any homogeneous state $\bar{x} \in X^{\mathcal{N}}_{\mathrm{diag}}$

is a fixed point of the interaction operator, $\mathcal{I}^{(\mathcal{N})}\bar{x} = \bar{x}$, i.e., interactions cannot change local coordinates if they are all equal. This property is a substitute of translation invariance for a multicomponent dynamical system with a finite number of local components.

If a number of local components in a spatially homogeneous system is infinite then there is a natural action $S^{(\mathcal{N})}$ of a (semi)group of (spatial) translations generated by shifts on some vector of the lattice formed by local systems. Therefore $S^{(\mathcal{N})} : X^{\mathcal{N}} \to X^{\mathcal{N}}$ can be viewed as a dynamical systems generated by spatial shifts. This group commutes in the spatially homogeneous case with time shifts (dynamics) and therefore the pair $(T^{(\mathcal{N})}, S^{(\mathcal{N})})$ forms a space-time dynamics.

The Definition 2.1 of a multicomponent dynamical system looks somewhat restrictive because of the assumption of superposition of a local dynamics and (spatial) interactions. However, it is not the case. Indeed, any map $\mathcal{T} : X^{\mathcal{N}} \to X^{\mathcal{N}}$ can be represented as $\mathcal{T} \equiv \mathcal{T} \circ (\bigotimes_{i\in\mathcal{N}} Id)$, where Id is an identity. In this representation the map $\mathcal{T}$ corresponds to the 'interaction' $\mathcal{I}^{(\mathcal{N})}$. However, this (formal) description while being relevant for multicomponent dynamical systems is not at all satisfactory for systems of interacting particles, which are the major models of interest in statistical mechanics. In these models the dynamics of individual particles (without interactions) is rather simple and often just trivial. Therefore it is the interactions that make the dynamics of interacting particles interesting. On the contrary, in multicomponent dynamical systems the dynamics of individual components (local systems) can be quite nontrivial (in fact, it could be any dynamical system). Therefore the property that interactions are just the identity on the diagonal is not an acceptable condition for statistical mechanics.

Perhaps the most popular example of interactions in multicomponent dynamical systems is a (spatially) homogeneous finite range coupling

$$(\mathcal{I}_\epsilon \bar{x})_i := (1-\epsilon)x_i + \epsilon \sum_{j=-k}^{k} a_j x_{i+j} \tag{2}$$

where the parameter ϵ is the strength of interactions, k is the radius of interactions, $a_i \geq 0$ are some constants, and $\sum_{i=-k}^{k} a_i = 1$. Observe that the set of all interactions described by (2) forms a convex hull of values of coordinates $\bar{x}$ in the k-neighborhood of the ith coordinate.

In the literature lattices of (interacting) maps often are referred to as CML or Lattice Dynamical Systems (LDS) if their time evolution can be or cannot be represented as the composition of a local dynamics and of (spatial) interactions respectively. Especially there is ubiquity of important examples which arise in various applications and are obtained by space discretizations of PDE where a multicomponent dynamical system has the form

$$(\mathcal{T}_\epsilon \bar{x})_i := (1-\epsilon)T_i x_i + \epsilon \sum_{j=-k}^{k} a_j x_{i+j} \tag{3}$$

i.e., an operator of interactions acts in the space $X^{\mathcal{N}}$ rather than on its shift $T^{(\mathcal{N})}X^{\mathcal{N}}$ under an (independent) local dynamics.

One cannot decompose formally this system into the superposition of the local dynamics (described by the map T_i) and spatial interactions $\mathcal{I}_\epsilon$. It is possible, however, to construct an equivalent system that satisfies both Definition 2.1 and the condition of invariance of the diagonal. The trick is to consider a replica of the local systems. Namely, for each local component we take its "delayed" copy acting on the phase space $Y_i \equiv X_i$ such that the interaction map Φ, $\Phi(\bar{x}, \bar{y}) := \bar{x}'$ is defined via the relation

$$x_i' = (1-\epsilon)x_i + \epsilon \sum_{j=-k}^{k} a_j y_{i+j} \ . \tag{4}$$

Then the local dynamics is defined as $\mathcal{T} : x_i \to (T_i x_i, x_i)$. It is easy to see that the multicomponent dynamical system $(\Phi \circ \mathcal{T}, \otimes_{i\in\mathcal{N}}(X_i \otimes Y_i))$ where the dynamics is the composition of a local dynamics and interactions onto x-component is the initial multicomponent system $(\mathcal{T}_\epsilon, \otimes_{i\in\mathcal{N}} X_i)$.

An important example of (2) is the diffusive coupling

$$(\mathcal{I}_\epsilon \bar{x})_i = (1-\epsilon)x_i + \frac{\epsilon}{3}(x_{i-1} + x_i + x_{i+1}) \tag{5}$$

i.e., the spatially homogeneous finite range interactions with $k = 1$ and $a_i = 1/3$. This system represents a discretization of the Laplacian. In fact,

$$-\epsilon x_i + \frac{\epsilon}{3}(x_{i-1} + x_i + x_{i+1}) = \frac{\epsilon}{3}((x_{i+1} - x_i) - (x_i - x_{i-1})) = \frac{\epsilon}{3}(\nabla x)_i.$$

Multicomponent dynamical systems (MDS) occupy an intermediate place between PDE and deterministic cellular automata (DCA). This is illustrated in the following table.

	Time	Space	Local (phase) Space
PDE	C	C	C
MDS	D	D	C
DCA	D	D	D

where C stands for "continuous" and D for "discrete."

It is worthwhile to mention that time in MDS can be continuous as well. We then just get coupled ordinary differential equations instead of coupled maps. A principal restriction is that the space (structure of a network of interacting local dynamical systems) in MDS is discrete. Indeed, although some of these models are obtained by discretization of PDE in the limit when the step of such discretization tends to zero, a character of dynamics of MDS usually changes uncontrollably. In fact, to keep the same type of dynamical behavior one needs to rescale the parameter(s) of the system in a peculiar

way (i.e., depending on an exact system we are dealing with). This lack of a limit transition to the system with continuous space is similar to the lack of transition from the discrete models of statistical mechanics to continuous models.

3 Deterministic (DCA) and Probabilistic (PCA) Cellular Automata

The dynamics of MDS is much richer than the dynamics of DCA because a local dynamics in DCA acts on a discrete local phase space and therefore it cannot be complex. On the other hand, a local dynamics in MDS can be quite complex (e.g., chaotic) and mimic random processes. Therefore the dynamics of MDS has, in a sense, the same complexity as the dynamics of PCA, where discreteness of a local phase space is "compensated" by random interactions between their (local) components. We consider for the sake of illustration and comparison with MDS two examples of CA, one DCA and one PCA. Both these examples belong to a popular class of CA of the "majority" type (or voter models).

Let x_i, $i \in \mathbb{Z}$ be a local variable assuming two values 1 and -1. The dynamics of a DCA is given via the relation

$$x_i^{n+1} = \mathrm{sgn}(x_{i-1}^n + x_i^n + x_{i+1}^n)$$

where n is the time variable. It is clear that every isolated plus or minus changes sign at the next moment, but clusters of pluses or minuses never change. Therefore two states with alternating signs $(\cdots + - + - + - \cdots)$ form a period two orbit, while all sequences with alternating clusters of pluses and minuses with lengths greater than one form an uncountable set U of fixed points. In fact, it is easy to prove that the set U is an attractor for all orbits except the period two orbit. Thus the dynamics of this DCA of a voter type is very simple.

The next example is the celebrated PCA introduced by A. Toom [54]. We start with a general definition of a probabilistic cellular automaton.

Let G be a finite or a countable graph. Consider a function $v : G \to \{1, 2, \ldots, K\}$ defined on the set of all vertices of this graph (which we denote by the same letter G) and assuming a finite number $K < \infty$ of values. A neighborhood $O(g)$ of a vertex $g \in G$ is the union of vertices to which g is connected. We assume that G is locally finite, i.e., for each vertex $g \in G$ its neighborhood $O(g)$ consists of at most $L < \infty$ vertices of the graph G.

A collection $v(g)$, $g \in G$ defines an instant configuration of PCA. A dynamics of PCA is a dynamics of its configurations. For each vertex $g \in G$ and for each configuration $v(O(g))$ of its neighborhood transition probabilities $p(v(g), v(O(g)), i)$, $i \in \{1, 2, \ldots, K\}$ are defined which determine a state (a value $v(g)$) at the vertex g at the next moment of time.

In the simplest version of Toom's example of PCA a graph G is the two-dimensional lattice $\mathbb{Z}^2$, and each vertex can be in two states $+1$ and -1. A neighborhood $O(g)$ of a vertex $g \in \mathbb{Z}^2$ in this PCA consists of two vertices $N(g)$ and $NW(g)$ where $N(g)$ is the nearest neighbor of g in the North direction, and $NW(g)$ is the nearest neighbor of $N(g)$ in the West direction. Transition probabilities are defined by a variant of a majority rule. It says that probabilities of the state $v_+(g)$ at the next moment of time equal

$$P(v_+(g) = +1) = 1 - \frac{\delta}{2} \quad \text{if } \operatorname{sgn}(v(g) + v(N(g)) + v(NW(g))) = +1$$
$$P(v_+(g) = +1) = \frac{\delta}{2} \qquad \text{otherwise .} \tag{6}$$

If $\delta = 0$ then Toom's PCA becomes the DCA for which the configurations of all pluses and of all minuses are fixed points of the dynamics.

The dynamics of PCA is (in general or "on average") simpler than the dynamics of MDS and in particular (again, generally) it is simpler to analyze. It is a reason why in many attempts to construct MDS with some desired properties the idea is to mimic some (class of) PCA by MDS. It is well known that the dynamics of (chaotic) dynamical systems is as rich as the dynamics of stochastic processes. We will explain now following [44] a simple construction which allows to represent Markov chains as deterministic dynamical systems (in fact as piecewise linear maps).

Let T be a nonsingular map from the unit interval $[0, 1]$ into itself. Denote by $\Delta = \{\Delta_i\}_{i=1}^{W}$ a partition of $[0, 1]$ into disjoint intervals Δ_i and $W \leq \infty$. We will call the partition Δ a *special* partition for the map T if $\forall\ i \leq W$, T is a diffeomorphism on the inner part of Δ_i onto its image.

The map T will be called Markov if there exists a special partition Δ such that $T\Delta_i = \bigcup_{j \in I(i)} \Delta_j$ for each i, where $I(i)$ is some set of indices. Such partition is called a Markov partition. Hence we deal here with special Markov partitions.

Consider now the following set of (probabilistic) measures on $[0, 1]$

$$\mathcal{M}_u([0,1], \Delta) = \left\{ \mu \in \mathcal{M}([0,1]) : \frac{\mu(I)}{\mu(I')} = \frac{|I|}{|I'|}, \right.$$
$$\left. \forall\ i \text{ and } \forall \text{ intervals } I, I' \subset \Delta_i \right\}$$

where $|I|$ is the Lebesgue measure (length) of the interval I. According to its definition $\mathcal{M}_u([0, 1], \Delta)$ (where u stays for "uniform") is the collection of all probability distributions with constant densities on each interval $\Delta_i \subset \Delta$, i.e., the restriction of any measure from this set to an interval Δ_i is proportional to its length for any $i \leq W$.

Theorem 3.1. [44] *For any given transition matrix P of a Markov chain with W states, $W \leq \infty$, there exists a piecewise linear one-dimensional map*

T with a Markov partition $\hat{\Delta}$ such that the restriction of the induced map T^* to $\mathcal{M}_u([0,1],\hat{\Delta})$ is equivalent to the left action of P in the space of probability distributions.

We outline a proof of this simple statement. Take any partition Δ of $[0,1]$ into $W \le \infty$ subintervals. For any integer $i \le W$, let $I_i = \{i_j\}_j$ denote the set of indices such that $P_{ii_j} > 0$. Then the refined partition, consisting of intervals Δ_{ij} with lengths $|\Delta_{ij}| = P_{ii_j}|\Delta_j|$ is a special partition $\hat{\Delta}$ for a piecewise linear map $T : [0,1] \to [0,1]$ defined as follows: on each interval Δ_{ij} the map T is a linear map from this interval onto the interval Δ_j. The partition $\hat{\Delta}$ is Markov since for any pair of indices i, j the image of Δ_{ij} under the action of T is an interval from the original partition, and hence a union of intervals from the partition Δ'.

Let $\{\hat{\Delta}_k\}$ be an enumeration of $\hat{\Delta}$. The density of the corresponding measure $\mu \in \mathcal{M}_u([0,1],\hat{\Delta})$ is piecewise constant on intervals $\hat{\Delta}_k$. We associate with the measure μ a vector of $\hat{\rho}(\mu)$ with components $(\hat{\rho}(\mu))_k = \mu(\hat{\Delta}_k)$. The constructed map T is Markov and linear on each element $\hat{\Delta}_k$. Therefore $(\hat{\rho}(T^*\mu))_k = (\hat{\rho}(\mu)P)_k$ for any k.

Observe that if a number of states W of this Markov chain is finite and the number of positive elements in the ith row of the matrix P does not exceed K, then the number of elements in the special partition for the constructed map does not exceed WK.

It is worthwhile to mention that generally it is not possible to construct a continuous map representing a given Markov chain. Consider, for example a Markov chain with three states and the transition matrix

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix} .$$

Then the described construction gives the following map [44]:

$$Tx = \begin{cases} 2x, & \text{if } 0 \le x < \frac{1}{3} \\ 2x - \frac{1}{3}, & \text{if } \frac{1}{3} \le x < \frac{2}{3} \\ 2x - \frac{4}{3}, & \text{if } \frac{2}{3} \le x < \frac{5}{6} \\ 2x - 1, & \text{if } \frac{5}{6} \le x \le 1 \, . \end{cases}$$

One cannot rearrange these four intervals to make the corresponding map continuous (even though we effectively got just four intervals instead of six provided by the construction). To place the map T to a circle instead of the interval would not help either. Indeed, the structure of the transition matrix P implies that the interval(s) corresponding to at least one of the three states of the Markov chain must be mapped to the intervals corresponding to two other states. Therefore there will be a gap there.

A much more important and interesting example is provided by a deterministic model of a random walk on non-negative integers. Such random walk is defined by transition probabilities $p_{i,i-1}$, $p_{i,i+1}$, $p_{i,i}$ to move at one step to the left, to the right or remain in the current position, and by the probabilities $p_{0,0}$ and $p_{0,1}$ for the boundary $i = 0$. Transition probabilities for the corresponding one-dimensional map T are $p_{i,j} = |\Delta_j|/|T\Delta_i|$ for any pair of indices i, j, $|i - j| \leq 1$.

Although Markov chains with a countable number of states form a rich class of models this class does not include PCA. Let $P_M : \mathcal{M}(X_M) \to \mathcal{M}(X_M)$ be a transfer operator corresponding to a Markov chain M acting on the space X_M of its states. We say that this Markov chain is *equivalent* to a dynamical system (T, X) if there exists a subspace $\mathcal{M}_M(X)$ of the space of (probabilistic) measures on X and a homeomorphism $\pi : \mathcal{M}(X_M) \to \mathcal{M}_M(X)$ such that $\pi \circ P_M = T^* \circ \pi$. This notion of equivalence is a generalization of the equivalence of the restricted induced map to the left action of the transition matrix discussed in Theorem 3.1.

Theorem 3.2. [44] *For any PCA on a locally finite graph G there exists a (deterministic) dynamical system generated by a countable number of maps with dynamics equivalent to the dynamics of this PCA.*

To construct such (equivalent to PCA) dynamical system at each vertex $g \in G$ we define a collection of maps corresponding to all possible configurations of states at g and in its neighborhood $O(g)$. A particular choice of the map is determined by the configuration of these states.

A total number of configurations (of values $v(g)$ and $v(O(g))$ cannot exceed $K^L < \infty$ for any $g \in G$. Observe that for any configuration $v(O(g))$ the construction in the proof of Theorem 3.1 can be readily applied to define a piecewise linear map $T_{v(O(g))}$ from $[0, 1]$ to itself with Markov partition consisting of at most K^K intervals Δ_i. This map has the dynamics equivalent to the one of our CA (defined by transition probabilities). Finally consider a multicomponent dynamical system with the phase space $X = [0, 1]^G$ (a direct product of unit intervals). At each vertex $g \in G$ a finite number of one-dimensional maps $T_{v(g)}$ is defined. Finally, for any point $\bar{x} = (x_g) \in X$ its image is defined via the following procedure. Let $x_{g'} \in \Delta_{i(g')}$ for any $g' \in O(g)$. Then the value of the gth coordinate of $\bar{x}$ at the next moment of time is defined by the map $T_{v(O(g))}(x_g)$, which concludes the construction required in Theorem 3.2. A particular case of this construction was considered in [43].

By making use of the same argument as above one can construct a MDS with $K = \infty$ as well if for any $i \in \{1, 2, \ldots, K\}$ only a finite number of transition probabilities (in the definition of a cellular automaton) are positive. However, the graph G still must be locally finite.

4 Relevant Measures for Dynamical Systems

Before discussing the problem of phase transitions in multicomponent dynamical systems, it is necessary to understand which class(es) of measures for dynamical systems are of interest in this respect. In fact, even without a general agreement what (if anything) one should mean by phase transitions in dynamical systems (not necessarily in spatially extended ones) majority of authors admit that one must relate the phenomenon of phase transitions with a change of a number of some measures for corresponding dynamical systems when their parameters vary. However, there is no general agreement which class(es) of measures one should consider in this respect (perhaps besides the obvious fact that only invariant measures matter). Moreover, the existing literature on this problem is very confusing. Indeed, different authors use the same names for different objects, different names for the same objects, etc. This situation brings even more confusions into a sophisticated problem on phase transitions in dynamical systems.

We will consider only dynamical systems with discrete time. However, the case of continuous time does not require any changes. It is customary when talking about relevant (classes of) measures for dynamical systems to refer to the review paper by Eckmann and Ruelle [58] where two classes of measures were singled out as reasonable candidates for relevant measures. One of this classes goes back to Kolmogorov although the corresponding measures are usually called stochastically stable (or zero-noise limit) measures. Real (physical, chemical, etc.) systems usually evolve in the presence of some noise (because of uncontrollable fluctuations of an "environment," etc.). Therefore their evolution must be described by some stochastic (Markov) process rather than by a deterministic dynamical system. If the stationary measures of these random processes (indexed by an amplitude ϵ of the noise) converge as $\epsilon \to 0$ to an invariant measure of the dynamical system then this measure is called stochastically stable (zero-noise limit or Kolmogorov) measure. However such measures may not exist (e.g., if a dynamical system has attractors with riddled basins). Another measures singled out in [58] are called Sinai-Ruelle-Bowen (or SRB) measures. SRB-measures are usually defined for sufficiently smooth hyperbolic dynamical systems (e.g., C^2 diffeomorphisms f) and are characterized by following two properties

(i) f has a positive Lyapunov exponent almost everywhere with respect to a measure μ and
(ii) μ has absolutely continuous conditional measures on unstable manifolds. For uniformly hyperbolic (Anosov, axiom A, etc.) systems the zero-noise limit and SRB measures coincide.

Both these classes of measures belong to more general class of *physical measures* [58]. A measure μ is called a physical measure for a dynamical system $f : X \to X$ if there exists a positive Lebesgue measure set $A \subset X$ such that for any continuous function $\phi : X \to \mathbb{R}$

$$\frac{1}{n}\sum_{i=0}^{n-1}\phi(f^i x) \underset{n\to\infty}{\longrightarrow} \int \phi\, d\mu \tag{7}$$

for every $x \in A$. The term "physical" suggests that such measures should be "observable" in physical experiments. It is not quite clear yet whether a set of positive Lebesgue measure in fact ensures a 'physical' observability, but certainly the notion of physical measures is very reasonable and widely accepted.

We will follow here a slightly more general approach suggested in [44] where instead of the dynamics of observables one deals with the dynamics of measures. Let (X, ρ) be a compact metric space with a certain reference measure m on it. Often m is taken as Lebesgue measure but it can be another measure as well. The dynamics on X is defined by a nonsingular, with respect to the measure m, map T (i.e., $m(T^{-1}A) = 0$ whenever $m(A) = 0$). Consider the space $\mathcal{M}(X)$ of probabilistic measures on X equipped with the topology of weak convergence. The induced map $T^* : \mathcal{M}(X) \to \mathcal{M}(X)$ is defined as $T^*\mu(A) = \mu(T^{-1}A)$ for any $\mu \in \mathcal{M}(X)$ and any Borel set $A \subseteq X$.

A measure μ_T is called a *natural measure* for the map T if there exists an open subset $U \subset X$ such that for every $\mu \in \mathcal{M}(X)$ absolutely continuous with respect to the reference measure m and having its support in U

$$\frac{1}{n}\sum_{k=0}^{n-1} T^{*k}\mu \underset{n\to\infty}{\longrightarrow} \mu_T\ . \tag{8}$$

The relation (8) means that μ_T is a stable fixed point of the dynamics of absolutely continuous (initial) measures. The set U is called the basin of attraction for the measure μ_T. (A similar definition has been used in [41, 59].)

A dynamical system can have several natural measures. Indeed, a natural measure is just an attractor of the action of T in the space of measures on X. One of the advantages of this definition of a natural measure is that it can be equally applied to random systems (Markov chains) as well. Indeed, let T^* be the transfer operator of a Markov chain with the phase space X. Then T^* generates the dynamics of measures on $\mathcal{M}(X)$. In other words, it is a conjugate operator to the transition matrix of this Markov chain.

The next definition of a *sample measure* is quite similar to the definition of a physical measure. The only difference that we now consider pathwise convergence of sample measures, rather than convergence of orbits of the induced map. A sample measure μ_T^p (where p stands for "point") is a common limit (as $n \to \infty$) of $\frac{1}{n}\sum_{k=1}^{n} \delta_{T^n x}$ for almost all points $x \in U$, where δ_x stands for the Dirac measure at the point x.

Both these definitions are based on Gibbs' idea of construction of stationary measures, and are closely related to the Bogoliubov-Krylov approach in the theory of dynamical systems. In the literature these two (natural and sample) measures are often identified. However, as we will see the relations between these measures can be nontrivial.

SRB measures for a dynamical system $T : X \to X$ are denoted by μ_T^c (where c stands for "conditional"). The following statement establishes relations between μ_T, μ_T^p, and μ_T^c.

Theorem 4.1. [44]

(a) *If μ_T^p exists, then μ_T is also well defined and $\mu_T = \mu_T^p$.*
(b) *If μ_T is absolutely continuous with respect to the reference measure m, invariant and ergodic then μ_T^p is well defined and $\mu_T^p = \mu_T$; however, without ergodicity this might not hold.*
(c) *If μ_T^c is stable with respect to the dynamics of absolutely continuous (initial) measures then $\mu_T = \mu_T^c$.*
(d) *The existence of μ_T or μ_T^p does not imply the existence of μ_T^c.*

For the proof of this theorem see [44]. We just consider here some instructive examples.

It is somewhat surprising that even if the sample measure μ_T^p exists and is unique it might be non-ergodic, which is illustrated by the map

$$Tx = \begin{cases} \left(1 - \sin\left(\pi x - \dfrac{\pi}{2}\right)\right)/2 & \text{if } 0 < x < 1 \\ x & \text{if } x = 0 \text{ or } x = 1\ . \end{cases} \tag{9}$$

The locally maximal attractor for this map consists of two fixed points 0 and 1. For any (initial) point $x \in (0, 1)$

$$\frac{1}{n} \sum_{k=1}^{n-1} \delta_{T^n x} \longrightarrow \frac{1}{2} (\delta_0 + \delta_1) = \mu_T^p\ . \tag{10}$$

However, this μ_T^p is obviously non-ergodic. Analogously, the images under the dynamics of any absolutely continuous probabilistic measure converge to $\mu_T = \frac{1}{2}(\delta_0 + \delta_1)$. So, μ_T is also non-ergodic and $\mu_T = \mu_T^p$.

The condition in (b) that μ_T must be invariant may seem to be redundant. In fact, it is not needed if the map T is continuous. However, for a general nonsingular map μ_T might not be invariant. Consider, for instance, the following map $T : [0, 1] \to [0, 1]$

$$Tx = \begin{cases} x/2 & \text{if } 0 < x \leq 1 \\ 1 & \text{if } x = 0. \end{cases} \tag{11}$$

It is easy to see that any probabilistic measure on $[0, 1]$ converges under the action of the induced map T^* to the δ-measure at 0. However, T has no invariant measure.

The statement that the natural measure might not coincide with the sample measure μ_T^p may look suspicious. To realize that it can nevertheless happen, consider the following map [60]

$$Tx = \begin{cases} x + 4x^3 & \text{if} \quad 0 \le x < \dfrac{1}{2} \\ x - 4(1-x)^3 & \text{if} \quad \dfrac{1}{2} \le x \le 1 \, . \end{cases} \tag{12}$$

This map has two neutral fixed points at 0 and 1. For any probabilistic absolutely continuous measure μ the sequence $T^{n*}\mu$ weakly converges to $\mu_T = \frac{1}{2}(\delta_0 + \delta_1)$ as $n \to \infty$ [61], which is non-ergodic (as for the map (9)).

Let $I_A(x)$ be the indicator function of the set A, i.e., $I_A(x) = 1$ if $x \in A$ and 0 otherwise. It has been shown [60] that for any sufficiently small $\delta > 0$ and for almost every point $x \in [0,1]$ (with respect to Lebesgue measure) $\limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} I_{[0,\delta]}(T^k x) = 1$ and $\liminf_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} I_{[0,\delta]}(T^k x) = 0$. Therefore, a sample measure μ_T^p does not exist. Thus the existence of a natural measure does not even ensure that μ_T^p exists.

A contracting map with a single globally attracting fixed point has δ-measure at this point and this is the unique invariant measure. Obviously both μ_T and μ_T^p coincide with this measure, while the measure μ_T^c does not exist in this case. (Sometimes people argue that μ_T^c for this map still exists because a (nonexisting!) unstable manifold of the globally attracting fixed point can be viewed as an empty set, and therefore δ-measure at this point can be viewed as an absolutely continuous one on this empty set. I believe that such a formal approach is just not much useful but rather harmful. For instance, it implies that this (the most stable) dynamical system is mixing, i.e., strongly chaotic, and thus claims that there is no difference whatsoever between regular and chaotic dynamics.)

The assumption in (b) that the natural measure μ_T is absolutely continuous is rather restrictive. It holds however for some important classes of dynamical systems, e.g., for maps satisfying the Collet-Eckmann condition and for piecewise expanding maps. Actually the conditions in (b) cannot be made essentially weaker, which is illustrated e.g., by the example of an expansive homeomorphism satisfying the specification property [62].

Another remark is that an SRB measure may exist but be not finite. It happens already for the closest to uniformly hyperbolic systems almost Anosov diffeomorphisms, i.e., for diffeomorphisms which are uniformly hyperbolic everywhere besides in a finite number of points. Similar results hold for one-dimensional locally expanding maps with some neutral points (see e.g., [12] and references therein).

Finally, it is worthwhile to mention recent surveys concerning SRB and physical measures [63, 64].

5 Phase Transitions in Multicomponent Dynamical Systems

The notion of a natural measure is well defined for MDS with a finite number of (local) components, i.e., $\mathcal{N} < \infty$. Indeed, in this case $(X^{\mathcal{N}}, \rho^{\mathcal{N}}, m^{\mathcal{N}})$ is also

a finite-dimensional metric space with a certain reference measure. Therefore one can apply and use for MDS with a finite number of components all the already discussed definitions and results.

One of the problems for MDS with an infinite number of components is that any probabilistic measure absolutely continuous with respect to $m^{\mathcal{N}}$ must coincide with it. Therefore, for infinite-dimensional multicomponent dynamical systems, one needs to use another approach to initial measures.

Let $\mathcal{L} \subseteq \mathcal{N}$ be a subset of the set of indices $\mathcal{N}$. Denote by $\pi^*_{\mathcal{L}} : \mathcal{M}(X^{\mathcal{N}}) \to \mathcal{M}(X^{\mathcal{L}})$ the projection operator in the space of probabilistic measures defined as $\pi^*_{\mathcal{L}}\mu = \int \mu d(\otimes_{i\in(\mathcal{N}\setminus\mathcal{L})} m_i)$ for any measure $\mu \in \mathcal{M}(X^{\mathcal{N}})$. Since there is a natural embedding of spaces $\mathcal{M}(X^{\mathcal{L}})$ into $\mathcal{M}(X^{\mathcal{N}})$ one can choose a family of distances $\mathrm{dist} = \mathrm{dist}_{\mathcal{L}}$ on these spaces such that

$$\mathrm{dist}(\pi^*_{\mathcal{L}}\mu, \mu) \underset{|\mathcal{L}|\to|\mathcal{N}|}{\longrightarrow} 0 \tag{13}$$

for every $\mu \in \mathcal{M}(X^{\mathcal{N}})$.

We call a measure $\mu \in \mathcal{M}(X^{\mathcal{N}})$ *smooth* if its marginals $\mu_i = \pi_{(i)}\mu$ are absolutely continuous with respect to the reference measures m_i for any $i \in \mathcal{N}$. We define an infinite-dimensional generalization of the natural measure μ_T as a common limit of Cesaro means $(1/n)\sum_{k=0}^{n-1}(T^{(\mathcal{N})*})^n\mu$ of all smooth measures $\mu \in \mathcal{M}(X^{\mathcal{N}})$ supported on a direct product $\otimes_{i\in\mathcal{N}} U_i$ of some open sets $U_i \subseteq X_i$, $i \in \mathcal{N}$. The set $\otimes_{i\in\mathcal{N}} U_i$ plays the role of the basin of attraction for μ_T.

Consider a family of multicomponent dynamical systems (T_γ, X) depending on some parameter γ. We say that this family has a *phase transition* at $\gamma = \gamma_c$ if the number of natural measures changes when the parameter γ crosses the value γ_c that is called a critical parameter value.

Observe that this may happen in two different ways. Let for $\gamma < \gamma_c$ every finite-dimensional approximation $(T^{(\mathcal{L})}_\gamma, X^{\mathcal{L}})$ has a finite number N_γ of natural measures $\mu^\gamma_{\mathcal{L}}$ and N_γ does not depend on $\mathcal{L}$. Assume also that any natural measure $\mu^\gamma_{\mathcal{N}}$ of the complete system is a (weak) limit of measures $\mu^\gamma_{\mathcal{L}}$. Then a phase transition occurs if all finite dimensional dynamical systems $(T^{(\mathcal{L})}_\gamma, X^{\mathcal{L}})$ (where $|\mathcal{L}|$ is large enough) i.e., "close" to $|\mathcal{N}|$) have a phase transition as γ crosses the value γ_c, i.e., a number N_γ changes at γ_c, and the same happens in the complete system.

However, (in the infinite-dimensional case only, i.e., when $|\mathcal{N}| = \infty$) there is also another scenario. Namely, finite-dimensional approximations (subsystems) may not demonstrate any phase transition, but their limit points either are not natural measures of $T^{(\mathcal{N})}_\gamma$ at the value γ_c, or new natural measure(s) of the complete systems appear which are not limit points of natural measures of finite-dimensional subsystems.

Many authors assume that only the second scenario can be called a phase transition in MDS in analogy with the approach of the equilibrium statistical mechanics. We rather believe that this point of view is too restrictive because

the MDS form much richer class than the systems of statistical mechanics. Indeed, as has already been noted, individual subsystems ('particles') in statistical mechanics in the absence of interactions with another subsystems have a simple dynamics, while in MDS such dynamics can be very complex.

With a slight abuse of notations, we denote by $\mathcal{T}^{(\mathcal{L})}$ a $|\mathcal{L}|$-dimensional approximation of the infinite-dimensional map $\mathcal{T}^{(\mathcal{N})}$ for a given subset $\mathcal{L} \subset \mathcal{N}$ of the set of indices. To define such an approximation explicitly one needs to specify boundary conditions, i.e., we have to choose the states of the local systems in the remaining infinite-dimensional part of the phase space.

It can be done in various ways. The most common ones are the fixed boundary conditions, where the coordinates of the vector $(\bar{x})_j$ for $j \in (\mathcal{N}\backslash\mathcal{L})$, do not change in time, and periodic boundary conditions, where $\mathcal{L}$ is a finite set of consecutive integers.

The following simple example [44] shows that a choice of the boundary conditions can change even very rough characteristics of the dynamics. Let $X_i = [0,1]$, and $T_i x \equiv Tx := x - ax(x-1)^2$, $0 < a < 1$ for all $i \in \mathcal{N} = \mathbb{Z}^1$, and $\mathcal{I}_\epsilon$ is the diffusive nearest-neighbors interactions. For a given subset of indices $\mathcal{L}$ denote by $\mathcal{I}_\epsilon^{(\mathcal{L},y)}$ the finite-dimensional approximation of the interaction map with the boundary conditions fixed at the value y for all coordinates in $\mathcal{N}\backslash\mathcal{L}$. Then for each value of the amplitude ϵ, $0 \le \epsilon \le 1$, of spatial interactions the multicomponent system $(\mathcal{I}_\epsilon^{(\mathcal{L},0)} T^{\mathcal{L}}, [0,1]^{\mathcal{L}})$ has only one attractor (and thus only one ergodic natural measure). On the contrary, if $0 < \epsilon < 3a/(8+2a)$ a number of attractors of the system $(\mathcal{I}_\epsilon^{(\mathcal{L},1)} T^{\mathcal{L}}, [0,1]^{\mathcal{L}})$ varies between 2 and $2^{|\mathcal{L}|}$ (depending upon the structure of the set $\mathcal{L}$). However, this multicomponent system has only one attractor in case of periodic boundary conditions. The corresponding proofs can be found in [44].

We will be mainly concerned with the situation when for each finite subset of $\mathcal{L}$ of indices and for a certain choice of boundary conditions the corresponding finite-dimensional approximations $\mathcal{T}^{(\mathcal{L})}$ have only one natural measure, while the entire system has several natural measures. In the statistical mechanics there are numerous examples of multi-particles systems with such behavior. An important example is provided by Toom's voter model (6). In this PCA there is one natural measure for $\delta > \delta_c$ but two natural measures (supported on configurations with predominantly $+1$s or -1s respectively) for $\delta < \delta_c$. Recall that in this model each local component can be in two states ($+1$ or -1), and the time evolution is govern by random "interactions" of local components, i.e., a future state of the local subsystem is determined by the present state of a certain subset of its neighbors with a random error. Therefore, according to our approach (definition) this model is not a multicomponent dynamical system.

A deterministic version of this model has been constructed in [43], where the randomness of interactions in Toom's model has been substituted by a collection of operators of local interactions for each local subsystem. A choice of a particular operator of local interactions is determined by the symbolic states

(rather than by exact values of local variables) in the same set of neighbors as in the Toom's PCA.

Observe that the procedure described in the Sect. 3 allows to construct a deterministic one-dimensional Markov map corresponding to a PCA. In fact, the construction in [43] basically follows the same idea. However, the resulting system is, of course, again equivalent to PCA and, in fact, have a trivial local dynamics. Indeed, the resulting spatial interaction is just a choice of one of several given maps. The result of this choice depends on states of neighboring elements. Thus such choice is, in fact, a random trial although formulated in different terms. The construction in Sect. 3 shows how such a translation from the probabilistic language to a (pseudo) dynamical one can be done in a general setting. In the heart of any such translation is a substitution of uncertainty of an outcome of a random trial by uncertainty of a local state when, instead of an explicit value (number) such state is labelled by a set (an element of some partition).

I believe that the MDS form a richer class of systems than the lattice models of statistical mechanics, and therefore one should adopt a more general approach to the phenomenon of phase transitions in these systems than the one of statistical mechanics. Especially phase transitions (i.e., changes in a number of natural measures) in MDS with a finite number of components are of interest if one considers parameters of spatial interactions (rather than internal parameters of the dynamics of (local) components).

Observe that it is not a problem of perturbation theory because we are interested in the regimes of dynamics which are different from the dynamics of noninteracting local systems (components). So it is a more sophisticated problem. Some such scenarios has been found (see e.g., [65, 66]). Especially the one discussed in [65] seems to be quite robust and rather general. However much more work should be done in this area.

It is confirmed formally [38] that the idea of phase transition [13, 17] leading to a more organized spatio-temporal dynamics works (if one adopts a general approach to phase transitions in MDS). It is tempting, however, to demonstrate that in MDS there are phase transitions similar to the ones in lattice models of statistical mechanics, especially in the Ising model. A hunt for an Ising type phase transition in CML started with the influential paper by Miller and Huse [36] where they showed numerically such phase transition in $\mathbb{Z}^2$ with nearest-neighbors diffusive interaction of piecewise linear antisymmetric one-dimensional maps. One of interesting questions in this area is whether Ising-type phase transitions can occur in one-dimensional CML. It is a well known result in statistical mechanics that there are no phase transitions on $\mathbb{Z}^1$. However, in CML on $\mathbb{Z}^1$ due to nontrivial dynamics of local maps there is another (second) dimension generated by time shifts (dynamics). There are some promising numerical results in this respect for CML on $\mathbb{Z}^1$ with modified Miller-Huse local maps [37, 45]. An interesting observation made in these studies is that a critical point corresponds to a minimum (with respect to the amplitude of interactions) of the Lyapunov dimension [67] in

the corresponding CML. Therefore the phase transition occurs at the point where these CML become the "most organized."

We conclude this section with the formulation of several conditions which ensure that phase transitions in a MDS cannot occur.

Theorem 5.1. [44] *Suppose that for any finite subset $\mathcal{L} \subset \mathcal{N}$ there exists only one natural measure of the induced map $\mathcal{T}^{(\mathcal{L})^*} := \pi_{\mathcal{L}} \mathcal{T}^{(\mathcal{N})^*} \pi_{\mathcal{L}}$, and there exist functions ϕ, $\psi : \mathbb{R}^1 \to \mathbb{R}^1$ and a constant $C < \infty$ such that for any two smooth measures $\mu, \nu \in \mathcal{M}(X^{\mathcal{N}})$ and for any two finite subsets $\mathcal{L} \subset \mathcal{L}' \subset \mathcal{N}$ the following conditions hold*

$$\operatorname{dist}(\mathcal{T}^{(\mathcal{L})^*}\mu, \mathcal{T}^{(\mathcal{L})^*}\nu) < C \operatorname{dist}(\mu, \nu) \tag{14}$$

$$\operatorname{dist}(\mathcal{T}^{(\mathcal{L})^*}\mu, \pi^*_{\mathcal{L}} \mathcal{T}^{(\mathcal{L}')^*}\mu) < \psi(|\mathcal{L}|) \underset{|\mathcal{L}|\to\infty}{\longrightarrow} 0 \tag{15}$$

uniformly in μ,

$$\operatorname{dist}(\mathcal{T}^{(\mathcal{L})^{*n}}\mu, \mu_{\mathcal{L}}) < \phi(n) \underset{|\mathcal{L}|\to\infty}{\longrightarrow} 0 \tag{16}$$

uniformly in $\mu, \mathcal{L}$. Assume also that

$$\operatorname{dist}(\mathcal{T}^{(\mathcal{L})^*}\mu, \mathcal{T}^{(\mathcal{N})^*}\mu) \underset{|\mathcal{L}|\to|\mathcal{N}|}{\longrightarrow} 0 \tag{17}$$

for any $\mu \in \mathcal{M}(X^{\mathcal{N}})$.
Then the multicomponent dynamical system $(\mathcal{T}^{(\mathcal{N})}, X^{\mathcal{N}})$ has only one natural measure.

Observe that in this theorem we do not assume that the dynamics is decomposable into a local dynamics and (spatial) interactions. Therefore the local components (subsystems) in the MDS may be non-identical. If we make such assumption than a simpler version of theorem can be formulated.

Theorem 5.2. [44] *Let $X_i \equiv X$, $T_i \equiv T$ for all $i \in \mathcal{N}$, and the map T is nonsingular with respect to the reference measure $m_i \equiv m$. Suppose also that the interaction $\mathcal{I}$ is local, i.e., the value $(\mathcal{I}\bar{x})_i$ depends only upon a finite number of "neighboring" components of x_i in the vector $\bar{x}$. Then the statement of Theorem 5.1 remains valid if we preserve the assumption (16), drop (17) and instead of (14) and (15) use weaker assumptions*

$$\operatorname{dist}(\mathcal{I}^*\mu, \mathcal{I}^*\nu) < \text{const } \operatorname{dist}(\mu, \nu) \tag{18}$$

$$\operatorname{dist}(\mathcal{I}^{(\mathcal{L})*}\mu, \pi^*_{\mathcal{L}} \mathcal{I}^{(\mathcal{L}')*}\mu) < \psi(|\mathcal{L}|) \underset{|\mathcal{L}|\to\infty}{\longrightarrow} 0 \tag{19}$$

for any $\mathcal{L}' \supset \mathcal{L}$ and $\mu, \nu \in \mathcal{M}(X^{\mathcal{N}})$.

So far the uniqueness of natural measures in the limit when the system's size goes to infinity has been proved for several classes of (local) chaotic maps in the region of weak spatial interactions [13, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31]. It remains to be seen though whether or not it is the only region where MDS do not have phase transitions.

Concluding Remarks

Coupled Map Lattices proved to be an exciting and interesting object to study, which influenced development of the entire area of dynamics of infinite-dimensional and, especially, spatially extended systems. In particular it allowed to make exact some crucial but somewhat vague notions as space time chaos, intermittency and coherent structures. The first successes of the theory of CML generated high expectations about their applications for real systems. However, it did not happen despite many efforts.

A mathematical theory of CML, on the other hand, became a respected part of the general theory of dynamical systems. A major challenge now is to apply the new notions, ideas and approaches developed in the theory of CML to the real systems where identity (or almost identity) of the components does not hold (as e.g., in the synchronization theory). The infinite dimension (number of components) is also a rather artificial condition. It is our opinion that the crucial question "what (spatial) interactions can do to dynamics of local (noninteracting) systems" remains widely open. To address this the most important problem one must move out of the region of weak interactions.

References

1. K. Kaneko, *Period doubling of kink-antikink patterns, quasiperiodicity in antiferro-like structures and spatial intermittency in coupled map lattices*, Prog. Theor. Phys. **72** (1984) 480–486. 9
2. K. Kaneko, *Spatial period-doubling in open flow*, Phys. Lett. **A111** (1985) 321–325. 9
3. I. Waller and R. Kapral, *Spatial and temporal structure in systems of coupled nonlinear oscillators*, Phys. Rev. **A30** (1984) 2047–2055. 9
4. R. Kapral, *Pattern formation in two-dimensional arrays of coupled discrete time oscillators*, Phys. Rev. **A31** (1985) 3868–3879. 9
5. S. P. Kuznetsov, *On critical behavior of one-dimensional lattices*, Pis'ma J. Technich. Fiz. **9** (1983), 94–98 (in Russian). 9
6. S. P. Kuznetsov, *On model description of coupled dynamical systems near the transition point order-disorder*, Izv. VUZov-Fizika **27** (1984) 87–96 (in Russian). 9
7. J. P. Crutchfield and K. Kaneko, *Phenomenology of spatio-temporal chaos*. In: *Directions in Chaos* (ed. Hao Bai-lin), World Scientific, Singapore, 1987, 272–353. 9, 10
8. *Theory and Applications of Coupled Map Lattices* (ed. by K. Kaneko), John Wiley, Chichester, 1993. 9, 30
9. Focus Issue on Coupled Map Lattices (ed. by K. Kaneko), *Chaos* **2** (1993). 9
10. *Lattice Dynamics* (ed. by H. Chaté and M. Courbage), Physica D **103** (1997), 1–612. 9
11. A. N. Kolmogorov, *A new metric invariant of transient dynamical systems and automorphisms of Lebesgue spaces*, Dokl. Akad. Nauk USSR **119** (1958), 861–864. 9
12. K. Kaneko and I. Tsuda, *Chaos and Beyond*, Springer, Berlin, 2000. 10, 24

13. L. A. Bunimovich and Ya. G. Sinai, *Space-time chaos in coupled map lattices*, Nonlinearity **1** (1988) 491–519. 10, 11, 12, 27, 28
14. Ya. G. Sinai, *Gibbs measures in ergodic theory*, Russ. Math. Surveys **27** (1972), 21–69. 10
15. D. Ruelle, *Thermodynamic Formalism*, Addison-Wesley, London, 1978. 10
16. R. Bowen and D. Ruelle, *The ergodic theory of Axiom A flows*, Invent. Math. **29** (1975) 181–202. 10
17. L. A. Bunimovich and Ya. G. Sinai, *Statistical mechanics of coupled map lattices*. In [8]: pp. 169–190. 10, 11, 27
18. J. Bricmont and A. Kupiainen, *Coupled analytic maps*, Nonlinearity **8** (1995) 379–396. 28
19. J. Bricmont and A. Kupiainen, *High temperature expansions and dynamical systems*, Comm. Math. Phys. **178** (1996) 703–732. 11, 28
20. J. Bricmont and A. Kupiainen, *Infinite-dimensional SRB measures*, Phys. D. **103** (1997) 18–33. 28
21. V. Baladi, M. Degli Esposti, S. Isola, E. Järvenpää and A. Kupiainen, *The spectrum of weakly coupled map lattices*, J. Math. Pures Appl. **77** (1998) 539–584. 28
22. Ya. B. Pesin and Ya. G. Sinai, *Space-time chaos in chains of weakly interacting hyperbolic mappings*, Advances in Soviet Math. **3** (1991) 165–198. 28
23. M. Jiang and Ya. B. Pesin, *Equilibrium measures for coupled map lattices*, Comm. Math. Phys. **193** (1998) 675–711. 28
24. M. Jiang, *Equilibrium states for lattice models of hyperbolic type*, Nonlinearity **8** (1995) 631–659. 28
25. M. Jiang, *SRB measures for lattice dynamical systems*, J. Stat. Phys. **111** (2003) 863–902. 28
26. G. Keller and R. Zweimüller, *Unidirectionally coupled interval maps: Between dynamics and statistical mechanics*, Nonlinearity **15** (2002) 1–24. 28
27. D. L. Volevich, *Kinetics of coupled map lattices*, Nonlinearity **4** (1991) 37–45. 28
28. D. L. Volevich, *The SRB measure for a multi-dimensional lattice of interacting hyperbolic mappings*, Dokl. Akad. Nauk USSR **47** (1993) 117–121. 28
29. D. L. Volevich, *Construction of an analogue of SRB measure for a multidimensional lattice of interacting hyperbolic mappings*, Russ. Acad. Math. Sbornik **79** (1994) 347–363. 28
30. E. Järvenpää, *An SRB-measure for globally coupled circle maps*, Nonlinearity **10** (1997) 1435–1469. 28
31. E. Järvenpää and M. Järvenpää, *On the definition of SRB measures for coupled map lattices*, Comm. Math. Phys. **220** (2001) 109–143. 11, 28
32. L. A. Bunimovich, *Lattice dynamical systems*. In: *From Finite to Infinite Dimensional Dynamical Systems* (ed. by J. C. Robinson and P. A. Glendinning), Kluwer, Dordrecht, 2001. 11
33. P. Grassberger and T. Schreiber, *Phase transitions in coupled map lattices*, Physica D **50** (1991) 171–188. 12
34. S. Sakaguchi, *Phase transitions in coupled Bernoulli maps*, Progr. Theor. Phys. **80** (1992) 7–13.
35. H. Chaté and P. Manneville, *Collective behavior in spatially extended systems with local interactions and synchronous updating*, Progr. Theor. Phys. **87** (1992) 1–60.
36. J. Miller and D. A. Huse, *Macroscopic equilibrium from microscopic irreversibility in a chaotic coupled map lattice*, Phys. Rev. **F 48** (1993) 2528–2534. 27

37. C. Boldrighini, L. A. Bunimovich, G. Cosimi, S. Frigio and A. Pellegrinotti, *Ising type transitions in coupled map lattices*, J. Stat. Phys. **80** (1995) 1185–1205. 27
38. L. A. Bunimovich and E. Carlen, *On the problem of stability in lattice dynamical systems*, J. Diff. Eqs. **123** (1995) 213–229. 27
39. P. Marcq and H. Chaté, *Early-time critical dynamics in lattices of coupled chaotic maps*, Phys. Rev. E **57** (1998) 1591–1603.
40. M. L. Blank, *Generalized phase transitions in finite coupled map lattices*, Physica D **103** (1997) 34–50.
41. M. L. Blank, *Stability and Localization in Chaotic Dynamics*, Moscow, MCCME, 2001. 22
42. G. Keller and M. Künzle, *Transfer operators for coupled map lattices*, Erg. Th. and Dyn. Syst. **12** (1992) 297–318.
43. G. Gielis and R. S. MacKay, *Coupled map lattices with phase transitions*, Nonlinearity **13** (2000) 867–888. 20, 26, 27
44. M. L. Blank and L. A. Bunimovich, *Multicomponent dynamical systems: SRB measures and phase transitions*, Nonlinearity **16** (2003) 387–401. 18, 19, 20, 22, 23, 26, 28
45. C. Boldrighini, L. A. Bunimovich, G. Cosimi, S. Frigio and A. Pellegrinotti, *Ising-type and other transitions in one-dimensional coupled map lattices with sign symmetry*, J. Stat. Phys. **102** (2001) 1271–1283. 27
46. W. Just, *Globally coupled maps: phase transitions and synchronization*, Physica D **81** (1995) 317–340.
47. S. Lepri and W. Just, *Mean field theory of critical coupled map lattices*, J. Phys. A **31** (1998) 6175–6181.
48. W. Just, *Equilibrium phase transitions in coupled map lattices: a pedestrian approach*, J. Stat. Phys. **105** (2001) 133–142. 12
49. S. Aubry, *The concept of anti-integrability: definitions, theorems and applications*. In *Twist Mappings and Their Applications* (ed. by R. McGehee and K. E. Meyer), Springer, 1992, 7–54. 12
50. R. S. MacKay, *Dynamics of networks: features which persist from the uncoupled limit*. In *Stochastic and Spatial Structures of Dynamical Systems* (ed. by S. J. van Strien and S. M. Verduyn Lunel), North Holland, 1996, 81–104. 12
51. S. Aubry, *Breathers in nonlinear lattices: existence, linear stability and quantization*, Physica D **103** (1997) 201–250. 12
52. I. S. Aranson, V. S. Afraimovich and M. I. Rabinovich, *Stability of spatially homogeneous chaotic regimes in unidirectional chains*, Nonlinearity **3** (1990) 639–651. 13
53. V. S. Afraimovich and L. A. Bunimovich, *Simplest structures in coupled map lattices and their stability*, Random and Comp. Dynamics **1** (1993) 423–444. 13
54. V. S. Afraimovich and Ya. B. Pesin, *Traveling waves in lattice models of multidimensional and multi-component media*, Nonlinearity **6** (1993) 429–455. 13, 17
55. V. S. Afraimovich and V. I. Nekorkin, *Chaos of traveling waves in a discrete chain of diffusively coupled maps*, Int. J. of Bif. and Chaos **4** (1994) 631–637. 13
56. A. Pikovsky, M. Rosenblum and J. Kurts, *Synchronization: A Universal Concept in Nonlinear Sciences*, Cambridge Univ. Press, 2002. 13
57. A. L. Toom, *Stable and attractive trajectories in multicomponent systems*. In: *Multicomponent Random Systems* (ed. by R. L. Dobrushin and Ya. G. Sinai), Dekker, NY, 1980.
58. J.-P. Eckmann and D. Ruelle, *Ergodic theory of chaos and strange attractors*, Rev. Mod. Phys. **57** (1985) 617–656. 21

59. D. J. Farmer, E. Ott and J. A. Yorke, *The dimension of chaotic attractors*, Physica D **7** (1983) 153–180. 22
60. T. Inoue, *Sojourn times in small neighborhoods of indifferent fixed points of one-dimensional dynamical systems*, Erg. Th. Dyn. Sys. **20** (2000) 241–258. 23, 24
61. G. Keller, *Completely mixing maps without limit measure*, Colloq. Math. **100** (2004), no. 1, 73–76. 24
62. M. Misiurewicz and A. Zdunik, *Convergence of images of certain measures.* In *Statistical Physics and Dynamical Systems* (ed. by J. Fritz, A. Jaffe and D. Szasz), Progr. Phys. Vol. 10, Birkhäuser, Boston, 1985, 203–219. 24
63. L.-S. Young, *What are SRB measures and which dynamical systems have them*, J. Stat. Phys. **108** (2002) 733–754. 24
64. M. Viana, *Dynamical systems: moving into the next century.* In: *Mathematics Unlimited:* 2001 *and Beyond*, Springer-Verlag, Berlin, 2001, 1167–1178. 24
65. L. A. Bunimovich and S. Venkatagiri, *Onset of chaos in coupled map lattices via the peak-crossing bifurcation*, Nonlinearity **9** (1996) 1281–1296. 27
66. L. A. Bunimovich and D. Turaev, *Localized solutions in lattice systems and bifurcations caused by spatial interactions*, Nonlinearity **11** (1998) 1539–1545. 27
67. P. Frederickson, J. L. Kaplan, E. D. Yorke, and J. A. Yorke, *The Lyapunov dimension of strange attractors*, J. Diff. Eqs. **49** (1983) 185–207. 27

On Phase Transitions in Coupled Map Lattices

W. Just[1] and F. Schmüser[2]

[1] MRC, Queen Mary / University of London, London, UK
w.just@qmul.ac.uk
[2] IAKS, University of Karlsruhe, Karlsruhe, Germany
schmues@ira.uka.de

Coupled map lattices are a paradigm for studying fundamental questions in spatially extended dynamical systems. Within this tutorial we focus on qualitative changes of the motion which are intimately related with the limit of large system size. Similar to equilibrium phase transitions, such qualitative changes are an ubiquitous feature of dynamical systems with a large number of degrees of freedom. Within the first section of this chapter we present an overview and some phenomenological facts of phase transitions in coupled map lattices. The following two sections describe in some details analytical tools which are useful for understanding phase transition behaviour in dynamical systems beyond plain numerical simulations. In Sect. 2 we explain how coupled map lattices are linked with the canonical equilibrium physics of spin systems when techniques of symbolic dynamics are applied. Using a simple model we explain how coupled map lattices are linked with phase transitions in equilibrium spin models. In the third section we describe an alternative approach in terms of kinetic spin models linking the dynamics of coupled map lattices with equilibrium and nonequilibrium statistical mechanics. We keep our presentation throughout this tutorial entirely elementary and confine the presentation to some basic concepts which are useful for tackling the analysis of phase transitions in extended dynamical systems.

1 Introduction and Overview

The dynamics of spatially extended dynamical systems is a long standing issue in theoretical physics and applied mathematical sciences. Prominent examples with strong emphasis on experimental aspects concern pattern formation out of equilibrium [1]. Due to the boost in the understanding of low-dimensional chaotic motion during the last two decades the emphasis has shifted recently towards the investigation of spatio–temporal dynamics. Unfortunately most of the underlying physical equations of motion like Navier–Stokes or Maxwell–Bloch equations are far too complex for analysing fundamental features

W. Just and F. Schmüser: *On Phase Transitions in Coupled Map Lattices*, Lect. Notes Phys. **671**, 33–64 (2005)
www.springerlink.com

beyond straightforward numerical simulations. Since time discrete models have proven to be fruitful in understanding low-dimensional chaos, coupled maps have become a paradigm for the investigation of space-time chaos [2, 3]. While the derivation of time discrete models may be based in principle on Poincaré cross sections such an approach does not apply for models with a spatially continuous variable, i.e. partial differential equations. Thus there still does not exist a satisfactory derivation of coupled map lattices from first principles[1].

Qualitative changes of the dynamical behaviour, i.e. topological changes of the phase space portrait are important and universal features. These changes are called bifurcations [4]. In spatially extended dynamical systems an additional mechanism may cause changes of the motion if the limit of a large number of degrees of freedom is considered. Such transitions share common aspects with equilibrium phase transitions in thermodynamic systems. In any finite canonical equilibrium ensemble the partition sum depends analytically on the parameters, e.g. the temperature. Non–analytic behaviour, i.e. qualitative changes may arise when the limit of infinite system size is considered. In fact, an approximate formal description of such equilibrium phase transitions in terms of a simple Ginzburg-Landau theory results in an order parameter equation which displays plain bifurcations (e.g. a pitchfork bifurcation for a homogeneous Ising system). Thus it became to some extent popular to term any type of bifurcation a phase transition (cf. e.g. [5]). Here we pursue a different nomenclature. While bifurcations may occur in any type of dynamical system, regardless of the number of degrees of freedom, there are changes of the dynamics which are absent in any finite size system and which occur only if the limit of infinite system size is considered[2]. In accordance with equilibrium thermodynamics we will call such qualitative changes (proper) phase transitions in order to distinguish these features from plain bifurcations. A formal definition of such a concept may become cumbersome (cf. [6]) and we refrain from giving a mathematical account. However, if one wants to study such aspects from a theoretical perspective the investigation of coupled map lattices is quite useful.

Coupled map lattices are spatio–temporal discrete dynamical systems. Consider a lattice, e.g. a spatially one-dimensional array of sites $\nu \in \{0, 1, \ldots, L-1\}$. On each lattice site consider a continuous variable taken e.g. from an interval $x^{(\nu)} \in I$. Then the state of the system is determined by an L-dimensional vector $\underline{x} = (x^{(0)}, \ldots, x^{(L-1)})$. As for the dynamics on each lattice site a map $f : I \to I$ acts on the corresponding component and the value at

[1] Sometimes one claims that numerical integration schemes for partial differential equations yield coupled map lattices. We find such an argument superficial. While integration schemes employ a small spacing between grid points, the lattice constant of coupled map lattices is usually large compared with the observed spatial structures. Thus both schemes focus on different asymptotic regimes.

[2] That means the long time limit and the limit of infinite system size do not commute.

time $n+1$ is determined from the state at time $n \in \mathbb{N}_0$ through the action of the map f and a coupling between neighbouring sites, e.g.

$$x_{n+1}^{(\nu)} = (1-\varepsilon) f\left(x_n^{(\nu)}\right) + \frac{\varepsilon}{2}\left[f\left(x_n^{(\nu-1)}\right) + f\left(x_n^{(\nu+1)}\right)\right] \tag{1a}$$

$$= \Phi_\varepsilon\left[f\left(x_n^{(\nu)}\right), f\left(x_n^{(\nu+1)}\right), f\left(x_n^{(\nu-1)}\right), \ldots\right] \tag{1b}$$

$$= \left(T_\varepsilon(\underline{x}_n)\right)^{(\nu)} \ . \tag{1c}$$

The coupling shown in (1a) is usually called diffusive coupling. There is no special reason for this choice apart from its simplicity and the hope that numerical features are independent from the particular form of the coupling (cf. [7] for a phenomenological analysis of particular models). The parameter $\varepsilon \in [0,1]$ indicates the strength of the coupling. The factor $(1-\varepsilon)$ is introduced for convenience as it ensures that the dynamics of the coupled map lattice (1a) is confined to the phase space I^L. Equation (1a) clearly indicates that coupled map lattices may be considered as the composition of two actions, namely the action of the single site map f on each lattice site and a coupling function $\Phi_\varepsilon : I^L \to I$ mediating the interaction as indicated in (1b). Such a formulation also indicates that coupling functions are of course not limited to the case of a diffusive (i.e. nearest neighbour) coupling and that quite different choices are possible. Finally, on an even more abstract level one may consider coupled map lattices just as dynamical systems acting on the phase space I^L, where the L-dimensional map T_ε depends on a coupling parameter in such a way that the choice $\varepsilon = 0$ yields for each component the action of the single site map f (cf. (1c)). Which type of formulation is preferred depends on the particular context. In some cases (e.g. when (1a) or (1b) are used) one needs in addition the specification of boundary conditions when starting the investigation for finite lattice size. Throughout our exposition we will use periodic boundary conditions $x^{(\nu)} = x^{(\nu+L)}$ since they are quite easy to handle. However, to the best of our knowledge no systematic investigation of the influence of boundary conditions is available in the literature.

The formulation of coupled map lattices in terms of (1b) or (1c) is quite flexible. Generalisations to spatial lattices of dimension larger than one are obvious, by taking the spatial index ν to enumerate the lattice sites and choosing appropriate coupling functions Φ_ε. There is no need to restrict the dynamical variables $x^{(\nu)}$ to intervals and to confine the full phase space to simple cubes I^L. It is even possible to start from the very beginning with a system on infinite lattices e.g. by considering $\nu \in \mathbb{Z}$. Such an approach, usually used in the mathematical context [6, 8, 9, 10, 11] is advantageous for performing rigorous proofs, although it calls for a careful definition of the dynamical system. Our strategy avoids these technicalities and starts from systems of finite size. The limit $L \to \infty$ will be considered when dynamical quantities have been computed. Thus our approach adopts a physicist's point of view of equilibrium phase transitions. In addition starting with a system of large but finite size directly refers to numerical simulations of space-time chaotic dynamics.

Finally the formulation in terms of (1b) or (1c) emphasises the role of different symmetries in the dynamical systems. Translation invariance is the most prominent one and we restrict our analysis to the translation invariant case. The study of dynamical systems with broken translation invariance, e.g. random coupled maps, is a science in its own (cf. e.g. [12] and references therein for aspects linking dynamical systems with glass transitions) and is still at its infancy.

Phase transitions in coupled map lattices are intimately linked with the limit of large system size. The effect of the system size on the dynamics can be demonstrated in terms of numerical simulations. A particularly simple example is given by dynamical systems with global coupling, i.e. when the coupling preserves permutation symmetry. A corresponding coupling function reads

$$\Phi_\varepsilon \left[x^{(\nu)}, x^{(\nu+1)}, x^{(\nu-1)}, \ldots \right] = x^{(\nu)} + \frac{\varepsilon}{L} \sum_{\mu=0}^{L-1} g\left(x^{(\mu)} - x^{(\nu)} \right) \tag{2}$$

where the analytical form of the interaction $g(x)$ depends on the particular model. The apparent simplicity of globally coupled models makes this class attractive for the investigation of fundamental properties of chaotic dynamical systems in the limit of large system size [13]. In fact, globally interacting chaotic maps may display nontrivial correlations even for weak interaction [14], provided the single site map f is structurally unstable [15, 16, 17, 18]. The aspect which concerns us most is the appearance of a phenomenon in strongly coupled models [19] which mimics a phase transition like behaviour. When one considers (structurally stable) chaotic single site maps the weak coupling regime is governed by a "space-time chaotic" state, i.e. a regime where the dynamics at different sites is uncorrelated. At a critical coupling strength ε_c, which depends on the particular choice of the single site map and the interaction function, the synchronised solution, i.e. the state $x_n^{(\nu)} = x_n^{(\mu)}$ for all μ and ν, becomes linearly stable. That transition takes place in any system of finite size L and is just a plain bifurcation. The former "space-time chaotic" state remains as a chaotic saddle yielding a "space-time chaotic" transient dynamics. Such a scenario is in full accordance with features known from low dimensional chaotic systems. But the transient times increase when the system size is increased (cf. Fig. 1 for an example of globally coupled Bernoulli maps). The observed exponential dependence of the mean transient time on the system size, $T_{\varepsilon,L} \simeq \exp(\alpha_\varepsilon L)$, may be understood in terms of a very simple geometric argument [20]. The synchronised solution $x_n^{(\nu)} = x_n^{(\mu)}$ denotes the diagonal in the phase space I^L. For strong coupling $\varepsilon > \varepsilon_c$ this state is linearly stable and possesses a trapping region such that the dynamics tends towards the synchronised solution once the trapping region is entered[3].

[3] In formal terms the trapping region may be defined to be a simply connected open subset of the basin of attraction which contains the synchronised solution.

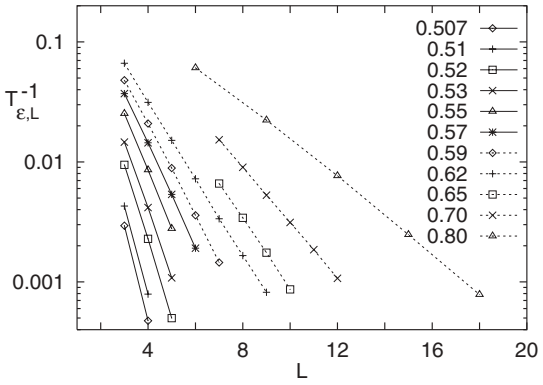


Fig. 1. Dependence of the inverse of the mean transient time (the relaxation rate) on the system size in a globally coupled map lattice for different values of the coupling strength. Confer [19] for further details of the model

From linear stability considerations the trapping region may be estimated by a cylinder with diameter d aligned along the diagonal in phase space. Its volume is estimated by Dd^{L-1} where D denotes the size of the domain I of the single site map f. In the phase space I^L there still exists the "space-time chaotic" saddle causing the chaotic transient dynamics. Assuming some degree of mixing a simple argument tells us that the probability for a typical initial condition to hit the trapping region is given by the volume ratio of the trapping region and the full phase space, Dd^{L-1}/D^L. This crude argument predicts the transient dynamics to be a Poisson process with probability $(d/D)^{L-1}$, i.e. with a time scale being given by $T \simeq (D/d)^L$ (cf. e.g. [19] for some numerical confirmation). Thus exponentially long transients may be explained in terms of simple geometric features, and such arguments may be applied quite generally at least on a phenomenological level. It is just such a property of the transient dynamics which is characteristic for phase transitions. Whereas no stable "nontrivial" dynamics persists in any system of finite size, since finally the synchronised solution is reached, the dynamics stays to be chaotic if systems of large size are considered. Even in models of moderate size L any simulation cannot overcome the transient regime of length $T \simeq \exp(\alpha_\varepsilon L)$. Thus to capture the relevant dynamics in terms of stationary states one has to consider first the limit of large system size and then the asymptotic limit of large time scales. Such a property is quite well known from equilibrium phase transitions[4]. Thus supertransients are an important indicator for the occurrence of phase transitions in spatio–temporal dynamics.

[4] One should however keep in mind that these arguments have to be considered with caution. There are types of transient behaviour, e.g. in diffusive systems, where time scales increase with system size as well, e.g. $T \simeq L^2$. That however usually does not cause phase transitions (cf. e.g. [21]). Thus the simple argument cannot replace a careful formal definition of phase transitions.

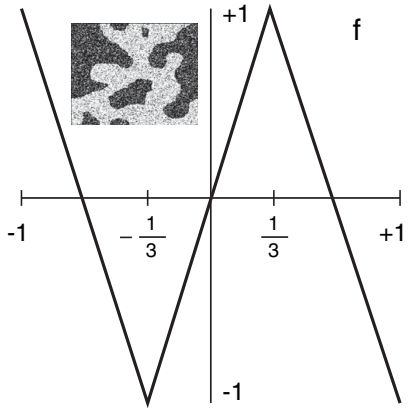


Fig. 2. Single site map of the Miller–Huse model. The inset shows a typical snapshot of pattern appearing in a spatially two-dimensional nearest neighbour coupled Miller–Huse model (*black/white* pixels encode the sign of the single site variable $x^{(\nu)}$). Confer [22] for further details of the model

A second class of quite popular examples of coupled map lattices displaying phase transitions are given by spatially two-dimensional coupled systems. The model introduced by Miller and Huse [22] is based on single site maps f of a double tent with an inversion symmetry (cf. Fig. 2). The single site dynamics is chaotic and develops random jumps for the sign $\sigma^{(\nu)} = \mathrm{sign}(x^{(\nu)})$ of the single site variable $x^{(\nu)}$. Thus such a type of motion can be considered as a kind of random spin dynamics, a so called kinetic spin model. Placing the maps on the sites of a two-dimensional square lattice and introducing a nearest neighbour coupling a phase transition appears which shares some common features with the two-dimensional equilibrium Ising model (cf. [23] for the original idea). In fact, numerical simulations of large systems display such a characteristic in a strong coupling regime. But a careful numerical analysis of the critical behaviour shows that naive universality arguments cannot be applied straightforwardly, i.e. the coupled map lattice inherits some hidden long range correlations yielding deviations from the Ising universality class [24, 25]. The description of coupled map lattices in terms of kinetic spin models on some coarse grained level looks promising and we will exploit the analytical structure of such approaches in detail in Sect. 3.

These two examples indicate what kind of feature one may expect when dealing with phase transitions in coupled map lattices. But solely on numerical grounds it seems difficult to judge whether a particular qualitative change of the dynamics is really related to the limit of infinite system size or whether the phenomenon is just caused by plain bifurcations (cf. e.g. the results in [3]). Therefore phase transitions in coupled map lattices call for formal analytical approaches although one might be forced to restrict the analysis to special model systems.

Already the seemingly simple case of weakly coupled maps poses a considerable challenge. When chaotic single site maps f are considered a naive view suggests that the weakly coupled system $\varepsilon \ll 1$ behaves to some extent like the uncoupled map, e.g. that spatio–temporal correlations decay exponentially and that the statistics of coordinates $x^{(\nu)}$ can be described by some distribution function. But the real proof of such conjectures is far from trivial and requires several technical assumptions concerning the map and the coupling (cf. e.g. [8, 9, 10, 11] for the analysis of various cases). In order to tackle such a problem one assigns symbols lattices to each spatio–temporal pattern. One dimension of the symbol lattice corresponds to the temporal evolution in the dynamical system whereas the other dimensions of the symbol lattice take the spatial extension of the dynamical system into account. Such approaches, called symbolic dynamics in a proper sense, are quite fruitful in linking the ergodic properties of dynamical systems to the statistics of symbol patterns, i.e. to the statistical mechanics of spin systems. In fact, such concepts have been developed in the context of low-dimensional chaos where the corresponding symbolic dynamics yields the statistical mechanics of symbol chains, i.e. one-dimensional spin systems. Then qualitative changes of the dynamics, i.e. bifurcations are reflected by equilibrium phase transitions caused by long range interactions (cf. e.g. [26] for the treatment of intermittent motion). We will consider the essence of such approaches in Sect. 2 using a simple example, so that most of the intricate technical difficulties which enter a more general discussion are avoided. Summarising these ideas, the dynamics of a d-dimensional coupled map lattice is mapped to the equilibrium statistical mechanics of a $d+1$ dimensional spin system. With certain technical assumptions one can show that the case of weakly coupled maps corresponds to the high–temperature phase of the spin system, so that all correlations, i.e. the spatio–temporal correlations in the map lattice, decay exponentially and the invariant measure is unique even in the limit of infinite system size. Increasing the coupling, i.e. decreasing the temperature in the associated spin system, may cause phase transitions when the limit of infinite system size is considered. Thus long range spatio–temporal correlations develop and the dynamical system possesses different ergodic components. One should however keep in mind that this type of approach targets at the limit of infinite system size. The structure of the dynamical system, e.g. the symbolic dynamics must be simple in order to apply the technique.

An alternative approach to deal with phase transitions in dynamical systems keeps the time as a dynamical variable and performs a coarse grained description with respect to the phase space variables only (cf. the discussion of the Miller–Huse model in the previous paragraphs). Based on partitions in phase space, e.g. by assigning at each time n symbols $\sigma_n^{(\nu)}$ according to the sign of $x_n^{(\nu)}$, one obtains a probabilistic model in the symbolic description where the transitions rates between different symbol states are determined by the underlying deterministic dynamics. In such a way one obtains descriptions

where on the level of the symbols a time evolution still exists, i.e. a kinetic Ising model or more general a probabilistic cellular automata. Such an idea can be worked out on a quite general level and may be applied to a variety of dynamical systems [27, 28]. But the focus here is on studying coupled map lattices. Application of such a program requires the knowledge of the temporal evolution of the phase space density. Thus one mostly deals with piecewise linear systems where the phase space densities are piecewise constant (cf. e.g. [29] for an investigation of one-dimensional maps or [16, 30] for the treatment of the mean field description of globally coupled maps). Using this property the just mentioned program can be performed explicitly and the dynamics of coupled map lattices may be studied beyond the weak coupling limit [6, 23]. The occurrence of phase transitions in short ranged coupled models seems to require an underlying spatially two-dimensional coupled map lattice in accordance with common wisdom of equilibrium statistical physics. We are devoting Sect. 3 for a more detailed discussion of such fruitful concepts.

There still does not exist a concise picture of phase transitions in spatially extended dynamical systems. At least coupled map lattices are a promising model class to study such a fundamental property. In what follows we will present methods and examples to cope with such a challenge.

2 Symbolic Dynamics of Coupled Maps

Symbolic dynamics is a tool to investigate the temporal evolution in terms of statistical properties of symbol arrays. Using a suitable partition of the phase space each initial condition generates an array of symbols according to the itinerary of the phase space point. Thus the dynamics is mapped to a symbol lattice where the time is translated into one spatial dimension of the lattice. In addition properties of the dynamical system are mapped to the equilibrium statistical mechanics of spin systems when an appropriate Hamiltonian is introduced. Although such approaches are nowadays contained in undergraduate textbooks (cf. e.g. [31, 32]) we try to keep our presentation self–contained and summarise the essential notations in Sect. 2.1. In fact we are not presenting a general theory here but just illustrate the essential features in terms of a simple model, a particular piecewise linear map. We will use this concept to demonstrate in Sect. 2.2 how the dynamics of coupled map lattices can be understood in terms of two-dimensional spin models. In particular, a coupled map lattice will be introduced such that the dynamics is equivalent to the nearest neighbour coupled Ising model. The nature of the phase transition and the consequences for the dynamics of the coupled map lattice is analysed in some detail. For our approach we rely for some technical reasons on the dynamics on invariant sets of saddle type, i.e. repellers. Thus we will in addition recall how such dynamical systems are related with iterated function schemes [33].

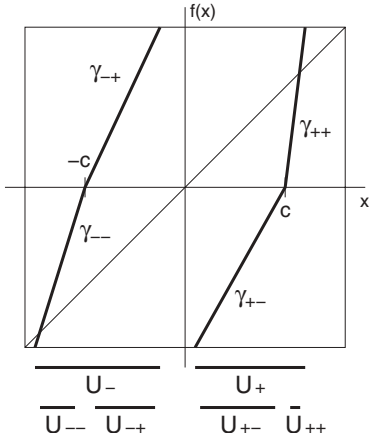


Fig. 3. Diagrammatic view of the map (3) and (4). The partition (5) and the first generation of cylinder sets is indicated as well

2.1 Basic Concepts for Piecewise Linear Maps

We now introduce some fundamental concepts in terms of a simple example, namely a one-dimensional map f which is piecewise linear on an interval, say $I = [-a, a]$ (cf. Fig. 3). In analytical terms the map is given by

$$f(x) = \begin{cases} g_-(x) & \text{if } x < 0 \\ g_+(x) & \text{if } x > 0 \end{cases} \tag{3}$$

where the two piecewise linear branches of the map are defined by

$$g_-(x) = \begin{cases} \gamma_{--}(x+c) & \text{if } x < -c \\ \gamma_{-+}(x+c) & \text{if } 0 > x > -c \end{cases} \tag{4a}$$

$$g_+(x) = \begin{cases} \gamma_{+-}(x-c) & \text{if } 0 < x < c \\ \gamma_{++}(x-c) & \text{if } x > c \end{cases} \tag{4b}$$

for some fixed value $0 < c < a$ and $\gamma_{\sigma\tilde{\sigma}} > 1$. The map is sketched in Fig. 3. By construction the map f maps points outside the interval I which are not considered any further, i. e. leave the domain of the map[5]. Points which stay in the interval I for one iteration are thus located in the two sets U_- and U_+ where

$$U_\sigma = g_\sigma^{-1}(I), \quad \sigma \in \{-, +\} \, . \tag{5}$$

Phase space points which stay within I upon two iteration steps are contained in one of the intervals $U_{\sigma_0\sigma_1} = g_{\sigma_0}^{-1}(U_{\sigma_1}) = g_{\sigma_0}^{-1}(g_{\sigma_1}^{-1}(I))$ where the symbol

[5] To ensure the structure depicted in Fig. 3 we suppose that the slopes of the map obey the inequalities $\gamma_{\pm\pm} > a/(a-c)$ and $\gamma_{\pm\mp} > a/c$.

string $\sigma_0\sigma_1$ tells us the itinerary of the phase space point with respect to the partition (5). Iterating this procedure we obtain the so called cylinder sets

$$U_{\sigma_0\sigma_1\ldots\sigma_{n-1}} = g_{\sigma_0}^{-1}(U_{\sigma_1\ldots\sigma_{n-1}}) = \{x \in I \,|\, f^k(x) \in U_{\sigma_k},\, 0 \le k \le n-1\}\,. \quad (6)$$

Taking "the limit" of large n one obtains the invariant set M of the map f, i.e. those phase space points which stay in I for an infinite number of iteration steps

$$M = \bigcap_{n\ge 1} \left(\bigcup_{\sigma_0\ldots\sigma_{n-1}} U_{\sigma_0\sigma_1\ldots\sigma_{n-1}} \right)\,. \quad (7)$$

Because of the expansiveness of the map f, $|f'(x)| > 1$, the length of the cylinder sets (6) decreases exponentially and the invariant set turns out to be a Cantor set with nontrivial fractal dimension. Since points which are not contained in M will leave the domain of the map our invariant set is of saddle type, i.e. a repelling invariant set.

The labelling of cylinder sets by symbol strings $\sigma_0\sigma_1\ldots\sigma_{n-1}$ has nice properties with respect to the dynamics. Just by definition (6) it follows that

$$f(U_{\sigma_0\sigma_1\ldots\sigma_{n-1}}) = g_{\sigma_0}(U_{\sigma_0\sigma_1\ldots\sigma_{n-1}}) = U_{\sigma_1\ldots\sigma_{n-1}} \quad (8)$$

and

$$f^{-1}(U_{\sigma_1\ldots\sigma_{n-1}}) = \bigcup_{\sigma_0} U_{\sigma_0\sigma_1\ldots\sigma_{n-1}} \quad (9)$$

while unions of cylinder sets obey

$$\bigcup_{\sigma_n} U_{\sigma_0\ldots\sigma_{n-1}\sigma_n} \subseteq U_{\sigma_0\ldots\sigma_{n-1}}\,. \quad (10)$$

Roughly speaking (8) tells us that the iteration of the map f corresponds to a symbol shift of a symbol string. Fixing an (infinite) symbol sequence $(\sigma_0, \sigma_1, \ldots)$ the corresponding sequence of cylinder sets U_{σ_0}, $U_{\sigma_0\sigma_1}$, $U_{\sigma_0\sigma_1\sigma_2}$, ... yields nested intervals which finally single out a unique point $x_{(\sigma_0,\sigma_1,\ldots)} \in M$, since the length of the cylinder sets decreases exponentially. In fact this recipe yields a homeomorphism (i.e. a continuous transformation with a continuous inverse) between the invariant set M and the space of all symbol sequences $\Sigma = \{-,+\}^{\mathbb{N}}$. Due to (8) this transformation nicely complies with the dynamics. Introducing the symbol shift S by $S : (\sigma_0, \sigma_1, \ldots) \mapsto (\sigma_1, \ldots)$ we obviously have $f(x_{(\sigma_0,\sigma_1,\ldots)}) = x_{S(\sigma_0,\sigma_1,\ldots)}$. This property is usually condensed in a commuting diagram[6]

[6] Since we are considering repelling sets our transformation is a conjugacy. Often such a property has to be relaxed. If one considers attracting sets then the relation between phase space points at the boundary of the partition and symbol sequences fails to be uniquely defined.

$$\begin{array}{ccc} S: \Sigma & \longrightarrow & \Sigma \\ \updownarrow & & \updownarrow \\ f: M & \longrightarrow & M \end{array} . \tag{11}$$

Thus the dynamics is equivalent to a symbol shift and orbits of the map f (i.e. initial conditions on the repeller) are efficiently labelled by symbol sequences. Our symbolic dynamics is based on the partition (5) of the phase space. Such partitions are usually called Markov partitions since the corresponding symbol sequences do not contain long ranged correlations between the symbols.

So far the construction has focussed on topological properties of the dynamics. We are now going to address the question "how frequently" a particular orbit, i.e. a symbol sequence, occurs. To be more definite we introduce probabilities for a symbol string $\sigma_0\sigma_1\ldots\sigma_{n-1}$ to occur when iterating the map f. Since the phase space points generating such a symbol string are contained in the corresponding cylinder set (6), the probabilities may be denoted by $\mu(U_{\sigma_0\sigma_1\ldots\sigma_{n-1}})$ where μ denotes an invariant measure. The condition of time invariance of the probabilities requires that (cf. (9))

$$\mu(U_{\sigma_1\ldots\sigma_{n-1}}) = \mu(f^{-1}(U_{\sigma_1\ldots\sigma_{n-1}})) = \sum_{\sigma_0} \mu(U_{\sigma_0\sigma_1\ldots\sigma_{n-1}}) \tag{12}$$

whereas the usual subadditivity of a measure yields (cf. (10)[7])

$$\mu(U_{\sigma_0\ldots\sigma_{n-1}}) = \sum_{\sigma_n} \mu(U_{\sigma_0\ldots\sigma_{n-1}\sigma_n}) . \tag{13}$$

Finally normalisation requires that

$$1 = \sum_{\sigma} \mu(U_\sigma) . \tag{14}$$

Actually (12), (13), and (14) just define the invariant measure, and there are many possible solutions of these equations. For instance if we take the ansatz

$$\mu(U_{\sigma_0\ldots\sigma_{n-1}}) = h_{\sigma_0\sigma_1} \frac{p_{\sigma_0\sigma_1}}{\alpha} \frac{p_{\sigma_1\sigma_2}}{\alpha} \cdots \frac{p_{\sigma_{n-3}\sigma_{n-2}}}{\alpha} \nu_{\sigma_{n-2}\sigma_{n-1}} \tag{15}$$

with nonnegative weights $p_{\sigma\tilde{\sigma}} \geq 0$ on the right-hand side, then the condition of invariance (12) results in

$$\alpha h_{\sigma_1\sigma_2} = \sum_{\sigma_0} h_{\sigma_0\sigma_1} p_{\sigma_0\sigma_1} \tag{16}$$

while the subadditivity (13) yields

$$\alpha \nu_{\sigma_0\sigma_1} = \sum_{\sigma_2} p_{\sigma_0\sigma_1} \nu_{\sigma_1\sigma_2} . \tag{17}$$

[7] Since the invariant measure is supported by M, (10) becomes an identity when the cylinder sets are restricted to M.

Equations (16) and (17) are nothing else but eigenvalue equations for the 4×4 matrix $A_{\sigma\tilde{\sigma},\tau\tilde{\tau}} = p_{\sigma\tilde{\sigma}}\delta_{\tilde{\sigma}\tau}$ which is usually called the transfer matrix of the map. Actually (16) is just the celebrated Frobenius–Perron equation [34] written with respect to a basis consisting of piecewise constant functions (cf. e.g. [35]). Finally normalisation (14) results in

$$1 = \sum_{\sigma_0\sigma_1} h_{\sigma_0\sigma_1}\nu_{\sigma_0\sigma_1} \ . \tag{18}$$

Thus given the weights $p_{\sigma\tilde{\sigma}}$ the parameters of the measure are determined[8].

We have computed a whole family of invariant measures where each measure is in fact ergodic. There still remains the problem which measure describes the dynamics of "typical" initial conditions. Since we are dealing with repelling sets, i.e. a Lebesgue typical initial condition leaves the domain of our map after a finite number of iterations, this question deserves a more detailed discussion. If one considers a uniform distribution of initial conditions in the interval I then after n iteration steps the fraction $\sum_{\sigma_0\ldots\sigma_{n-1}} \lambda(U_{\sigma_0\ldots\sigma_{n-1}})$ is still contained in I, where $\lambda(U_{\sigma_0\ldots\sigma_{n-1}})$ denotes the normalised length (i.e. the normalised Lebesgue measure) of the cylinder set. By construction (cf. (8)) the length obeys

$$\lambda(U_{\sigma_0\sigma\ldots\sigma_{n-1}}) = \frac{1}{\gamma_{\sigma_0\sigma_1}}\frac{1}{\gamma_{\sigma_1\sigma_2}}\cdots\frac{1}{\gamma_{\sigma_{n-2}\sigma_{n-1}}}\lambda(U_{\sigma_{n-1}}) \tag{19}$$

since in each iteration step the map f expands by a factor $\gamma_{\sigma\tilde{\sigma}}$. Using this property and some standard transfer matrix argument the fraction of initial conditions remaining in the domain of the map scales exponentially (cf. [36])

$$\sum_{\sigma_0\ldots\sigma_{n-1}} \lambda(U_{\sigma_0\ldots\sigma_{n-1}}) \simeq \alpha^n \tag{20}$$

where the (topological) escape factor $\alpha < 1$ obeys the eigenvalue (15) and (16) if the weights are identified with the inverse slopes

$$p_{\sigma\tilde{\sigma}} = \frac{1}{\gamma_{\sigma\tilde{\sigma}}} \ . \tag{21}$$

Thus the measure (15) determined by the inverse slopes of the map is apparently related with the dynamics of "typical" initial conditions. One may pursue this argument by introducing the concept of conditional invariant measures [37]. Consider a uniform distribution of initial conditions. The fraction $\mu_0^c(f^{-1}(I))$ of initial conditions stays in the interval I when performing an iteration step where $\mu_0^c = \lambda$ denotes the (normalised) Lebesgue measure. Therefore $\mu_1^c(E) = \mu_0^c(f^{-1}(E))/\mu_0^c(f^{-1}(I))$ denotes the probability that after one

[8] As usual a common factor between the left and the right eigenvector remains undetermined. However, such a fact does not influence the expression of the measure (15).

iteration step phase space points appear in the set $E \subseteq I$, provided those initial conditions leaving I are ignored. Iterating this prescription the quantity $\mu^c_{k+1}(E) = \mu^c_k(f^{-1}(E))/\mu^c_k(f^{-1}(I))$ yields the fraction of initial conditions, ending up in E after $k+1$ iteration steps. Using $\mu^c_0(f^{-1}(I)) = \sum_\sigma \lambda(U_\sigma)$, (9) and (19), straightforward iteration results in

$$\begin{aligned}\mu^c_{k+1}(U_{\sigma_0\dots\sigma_{n-1}}) &= \frac{\displaystyle\sum_{\sigma_{-k-1}\dots\sigma_{-1}} \mu^c_0(U_{\sigma_{-k-1}\dots\sigma_{-1}\sigma_0\dots\sigma_{n-1}})}{\displaystyle\sum_{\sigma_{-k}\dots\sigma_0} \mu^c_0(U_{\sigma_{-k}\dots\sigma_0})} \\ &= \frac{\displaystyle\sum_{\sigma_{-k-1}\dots\sigma_{-1}} \gamma^{-1}_{\sigma_{-k-1}\sigma_{-k}}\cdots\gamma^{-1}_{\sigma_{-1}\sigma_0}\lambda(U_{\sigma_0\dots\sigma_{n-1}})}{\displaystyle\sum_{\sigma_{-k}\dots\sigma_0} \gamma^{-1}_{\sigma_{-k}\sigma_{-k+1}}\cdots\gamma^{-1}_{\sigma_{-1}\sigma_0}\lambda(U_{\sigma_0})} \,. \end{aligned} \tag{22}$$

Standard arguments of linear algebra[9] tell us that the limit

$$h_{\sigma_0\sigma_1} = \lim_{k\to\infty} \frac{\displaystyle\sum_{\sigma_{-k-1}\dots\sigma_{-1}} \gamma^{-1}_{\sigma_{-k-1}\sigma_{-k}}\cdots\gamma^{-1}_{\sigma_{-1}\sigma_0}}{\displaystyle\sum_{\sigma_{-k}\dots\sigma_0} \gamma^{-1}_{\sigma_{-k}\sigma_{-k+1}}\cdots\gamma^{-1}_{\sigma_{-1}\sigma_0}\lambda(U_{\sigma_0})} \tag{23}$$

exists since the numerator behaves like $\alpha^k h_{\sigma_0\sigma_1}$ and the denominator like α^k where α denotes the largest eigenvalue and $h_{\sigma_0\sigma_1}$ the corresponding and properly normalised eigenvector of the eigenvalue (16) with the choice (21). Thus taking the limit $k \to \infty$ we obtain the stationary probabilities (the so called conditional invariant measure)

$$\mu^c_\infty(U_{\sigma_0\dots\sigma_{n-1}}) = h_{\sigma_0\sigma_1}\lambda(U_{\sigma_0\dots\sigma_{n-1}}) \,. \tag{24}$$

Equation (24) determines the probability distribution of phase space points when the density is properly rescaled after each iteration step, taking the loss of probability into account. The right-hand side tells us that such a probability is described in terms of a density function where the density is determined by the eigenvalue problem (16) with the special choice (21). In this sense the invariant measure (15) subjected to the choice (21) describes the distribution of typical initial conditions, while the adjoint eigenvalue problem (17) takes the fractal support of the measure into account [38]. Measures which describe the ergodic properties with respect to typical initial conditions are usually called SRB measures. In general they are linked with the local expansion rates as expressed by the fundamental relation (21). Henceforth we will restrict our

[9] Using e.g. the previously mentioned transfer matrix $A_{\sigma\tilde{\sigma},\tau\tilde{\tau}} = \gamma^{-1}_{\sigma\tilde{\sigma}}\delta_{\tau\tilde{\tau}}$ (cf. (16) and (17)), the numerator and denominator can be expressed in terms of the kth power of this transfer matrix.

analysis to this choice. There exists a whole multifractal industry analysing the implications of the other cases and we refer the interested reader to the literature [39, 40, 41].

Properties of the dynamical system, e.g. mean values and correlation functions can be evaluated in terms of the invariant measure (15). If we consider for the purpose of illustration the simplest case, i.e. a piecewise constant observable $G(x) = G_\sigma$ if $x \in U_\sigma$, then the mean value is given by

$$\begin{aligned}\langle G\rangle &= \int G(x)\, d\mu = \sum_\sigma G_\sigma \mu(U_\sigma) \\ &= \frac{1}{n}\sum_{k=0}^{n-1} \sum_{\sigma_0 \ldots \sigma_{n-1}} G_{\sigma_k} \mu(U_{\sigma_0 \ldots \sigma_{n-1}}) \end{aligned} \tag{25}$$

where for the last expression we have employed the conditions of invariance (12) and (13)[10]. For the temporal correlation function one obtains in the same way

$$\begin{aligned}\langle G(x)\; G(f^\ell(x))\rangle &= \int G(x) G(f^\ell(x))\, d\mu = \sum_{\sigma_0 \ldots \sigma_\ell} G_{\sigma_0} G_{\sigma_\ell} \mu(U_{\sigma_0 \ldots \sigma_\ell}) \\ &= \frac{1}{n}\sum_{k=0}^{n-1} \sum_{\sigma_0 \ldots \sigma_{n+\ell-1}} G_{\sigma_k} G_{\sigma_{k+\ell}} \mu(U_{\sigma_0 \ldots \sigma_{n+\ell-1}})\,. \end{aligned} \tag{26}$$

Inspecting (25) and (26) the average values can be expressed as canonical equilibrium averages of a single spin variable G_σ when one defines the spin Hamiltonian through the statistical weight

$$\mu(U_{\sigma_0 \ldots \sigma_{n-1}}) = \exp(-H_{\sigma_0 \ldots \sigma_{n-1}})\,. \tag{27}$$

The Hamiltonian $H_{\sigma_0 \ldots \sigma_{n-1}}$ defines the statistical mechanics of a spin chain where the spatial extension corresponds to the temporal evolution of the dynamical system. For our model (15) and (21) the Hamiltonian reads

$$\begin{aligned} H_{\sigma_0 \ldots \sigma_{n-1}} &= -\ln h_{\sigma_0 \sigma_1} - \ln \nu_{\sigma_{n-2}\sigma_{n-1}} + \sum_{k=0}^{n-2} \ln \gamma_{\sigma_k \sigma_{k+1}} + (n-2)\ln\alpha \\ &\simeq \sum_{k=0}^{n-1} (-J\sigma_k\sigma_{k+1} - h\sigma_k) + n\ln\alpha \end{aligned} \tag{28}$$

where $4J = -\ln[\gamma_{++}\gamma_{--}/(\gamma_{+-}\gamma_{-+})]$, $2h = -\ln[\gamma_{++}/\gamma_{--}]$, and we use the symbols σ_k as well to indicate the spin quantum number ± 1. Thus apart from boundary conditions which have been omitted in the last line of (28)

[10] Actually the last expression is useful if one wants to generalise the considerations to more complicated variables where the limit of a fine partition is really needed.

the Hamiltonian is given by the nearest neighbour coupled Ising chain. The exchange constant is determined by the curvature of the map, the external field by the symmetry, and the ground state energy by the escape rate.

So far the formulation of ergodic averages in terms of canonical ensembles is just a formal procedure. It does not help to compute the dynamical properties since the invariant measure must be known a priori in order to determine the Hamiltonian. But standard results of statistical mechanics tell us how the mean values and the correlations depend on the parameters of the Hamiltonian, i.e. on the parameters of the map. Such results are in particular valid in the thermodynamic limit $n \to \infty$, i.e. in the limit of fine partitions of the phase space. Since we have obtained a spin chain with short range interactions canonical averages are analytic expressions and the correlations decay exponentially, even if quite complicated observables $G(x)$ are considered. Thus we obtain that the mean value (25) depends analytically on the parameters of the map and that the correlation (26) decays exponentially. Summarising, by adopting a symbolic description and translating the time into a lattice dimension dynamical properties may be reformulated within the concepts of canonical equilibrium statistical mechanics. For expansive dynamical systems we obtain Hamiltonians with short range interaction. No (equilibrium) phase transitions occur and analytic dependence of the mean values on the parameters of the system is obtained. If expansiveness is violated then long range interactions are caused and phase transitions may occur. A prominent example is intermittent dynamics where the bifurcations in the dynamical system are mapped to particular phase transitions in the corresponding equilibrium statistical mechanics [26].

A direct numerical simulation of our model is not feasible since Lebesgue almost all initial conditions leave the domain I. But considering the "inverse" dynamics in terms of an iterated function scheme [33] we obtain a nice method to generate a time series and to compute the invariant density by numerical means. In fact such an approach just formalises the construction of the invariant set M, (7), by a sequence of cylinder sets. Starting with our initial domain $E_0 = I$ the preimage $E_1 = f^{-1}(E_0) = \cup_\sigma g_\sigma^{-1}(E_0)$ consists of the two intervals (5). We may iterate such a procedure $E_{k+1} = \cup_\sigma g_\sigma^{-1}(E_k)$ so that E_k consists of the union of cylinder sets of order k. Since our original map is expansive the inverse branches g_σ^{-1} are contractions and the sequence (E_k) converges towards the (maximal) invariant set M of our dynamical system. It is determined by the condition (cf. e.g. [33] for a nice introduction containing the technical details)

$$M = \bigcup_\sigma g_\sigma^{-1}(M) \; . \tag{29}$$

The collection of contractions $\{g_\sigma^{-1}\}$ is called an iterated function scheme. We may apply this idea to generate orbits of our dynamical system. Suppose the symbol sequence $(\sigma_0, \sigma_1, \ldots)$ of an orbit is given. Then the orbit obeys $x_{n+1} = f(x_n) = g_{\sigma_n}(x_n)$. Inversion of the branch yields the reversed dynamics

$x_n = g_{\sigma_n}^{-1}(x_{n+1})$. Since the inverse branches are contractions, any initial value converges after a short transient phase towards the true orbit. Of course in order to generate a "typical" orbit (i.e. to take a "typical" symbol sequence) one has to formulate the concept of an invariant measure using the iterated function scheme. By definition the invariant measure obeys

$$\mu(E) = \mu(f^{-1}(E)) = \bigcup_{\sigma} \mu(g_\sigma^{-1}(E)) \tag{30}$$

for any (measurable) set E. In order to generate a μ-typical symbol sequence we consider the conditional probability that a symbol σ_n appears provided that the symbols σ_k, $k \geq n+1$ are given

$$w(\sigma_n|\sigma_{n+1}, \sigma_{n+2}, \ldots) = \lim_{k\to\infty} \frac{\mu(U_{\sigma_n\sigma_{n+1}\ldots\sigma_k})}{\mu(U_{\sigma_{n+1}\ldots\sigma_k})} \,. \tag{31}$$

Choosing symbols according to these transition probabilities we generate a μ-typical orbit through

$$x_n = g_{\sigma_n}^{-1}(x_{n+1}) \quad \text{with probability } w(\sigma_n|\sigma_{n+1}, \sigma_{n+2}, \ldots) \tag{32}$$

where the symbols σ_k, $k \geq n+1$ are determined by x_k. Such a type of stochastic model yields a typical orbit on our repeller. For our particular case (15) the transition probabilities yield a Markov chain since evaluation of (31) results in

$$w(\sigma_n|\sigma_{n+1}, \sigma_{n+2}, \ldots) = \frac{h_{\sigma_n\sigma_{n+1}}}{h_{\sigma_{n+1}\sigma_{n+2}}} \frac{p_{\sigma_n\sigma_{n+1}}}{\alpha} \tag{33}$$

and the eigenvector $h_{\sigma_{n+1}\sigma_{n+2}}$ does not depend on the second symbol (cf. (16)). The dynamical system (32) admits a skew product structure. The stochastic part may be considered as a single spin flip dynamics with short term memory. The iteration rule for the phase space point x_n is then derived from the dynamics of the symbols.

The invariant measure may be computed in terms of a histogram from a typical trajectory. Since our dynamical system is supported by a fractal no smooth density exist and visualisation calls for a cumulative distribution, e.g. $P_<(x) = \mu((-\infty, x])$. If one considers the simplest case, i.e. linear contractions g_σ^{-1}, then the cumulative distribution clearly shows a devil's staircase like behaviour and thus reflects the Cantor set structure of the invariant set (cf. Fig. 4). Introducing a finite external field into the corresponding statistical mechanics breaks the symmetry of the map and the invariant measure. If we introduce a finite curvature in the branches of the map, i.e. if we take a finite nearest neighbour interaction in the statistical mechanics into account no qualitative change occur. However, in the low temperature limit, i.e. $J \to \infty$, a singular point is approached since the decay of equilibrium correlation functions of the nearest neighbour Ising model slows down. We may study how this behaviour is reflected by the dynamical system. There are in fact different ways approaching the asymptotic limit, depending on the behaviour of

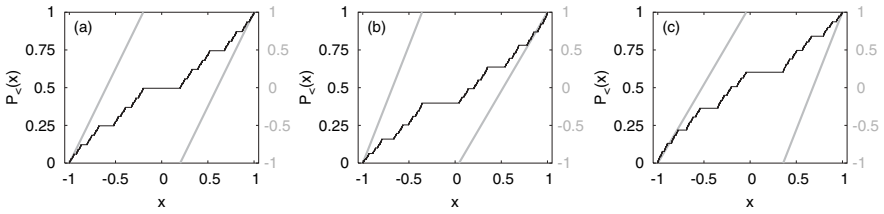


Fig. 4. Cumulative distribution $P_<(x)$ of a piecewise linear model with $J = 0$ and (**a**) $h = 0$, (**b**) $h = 0.2$, and (**c**) $h = -0.2$ (cf. (28)). In each case the escape factor is chosen to be $\alpha = 0.8$. The shape of the single site map f is sketched in *grey*

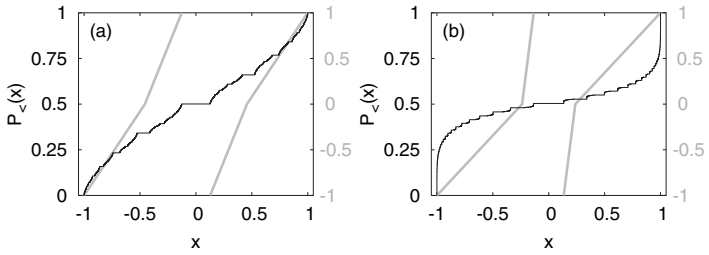


Fig. 5. Cumulative distribution $P_<(x)$ of a piecewise linear model with $h = 0$ and (**a**) $J = 0.25$, (**b**) $J = 1.0$ (cf. (28)). In each case the escape factor is chosen to be $\alpha = 0.8$. The shape of the single site map f is displayed in *grey*

the ground state energy, i.e. the escape factor. Figure 5 summarises some typical results. One clearly recognises that the cumulative distribution develops a square root like singularity at the fixed points of the map, $x = \pm 1$. Thus the invariant measure accumulates in this phase space region. The neighbourhood of the fixed point generates symbol sequences containing long strings of symbols with the same sign, i.e. the fixed points corresponds to the ground states of the Ising Hamiltonian. Furthermore since the fixed points tend to become marginally stable in the limit $J \to \infty$ the decay of temporal correlations slows down. Thus the behaviour is a caricature of the intermittency transition mentioned previously [26, 41]. The two pure phases correspond to the two fixed points of the map and the invariant measure tends to localise at the fixed points. Of course no phase transition is caused since we have considered a degenerated limit.

In summary, low-dimensional dynamical systems have been cast into the framework of canonical equilibrium statistical mechanics of spin chains. In this approach the spatial coordinate of the spin system corresponds to the temporal evolution of the dynamical system. Since the invariant measure defines the spin Hamiltonian, the explicit construction of the statistical mechanics requires the full solution of the dynamical model. One may generalise these concepts to more complicated dynamical systems, e.g. higher dimensional expanding maps [42]. The crucial step for the application of the present concept

is the knowledge of the underlying symbolic dynamics (cf. e.g. [43]). Since our focus is on fundamental aspects in coupled map lattices we stay with our simple model. For the statistical mechanics of general low-dimensional dynamical systems we refer the interested reader to the literature [44].

2.2 Coupled Repeller Maps

Based on the approach sketched in Sect. 2.1 we are going to introduce a coupled map lattice exhibiting an Ising–like phase transition. Our construction closely follows the elementary ideas of [45]. Consider a spatially one-dimensional lattice of length L and place on each lattice site the single site map (3). To introduce the essential notation let us first consider the trivial case without coupling, $\Phi_{\varepsilon=0}[x^{(\nu)},\ldots] = x^{(\nu)}$. Then the dynamics is governed by

$$x_{n+1}^{(\nu)} = (T_{\varepsilon=0}(\underline{x}_n))^{(\nu)} = f\left(x_n^{(\nu)}\right) \ . \tag{34}$$

On each lattice site a symbol $\sigma^{(\nu)} \in \{-,+\}$ indicates which branch (4) of the single site map is applied. Actually the uncoupled map (34) is not defined on the full phase space I^L since points leave the domain on iteration. The partition (5) of the single site map carries over to a partition of the spatially extended system in the trivial way

$$U_{\underline{\sigma}} = U_{\sigma^{(0)}} \times U_{\sigma^{(1)}} \times \cdots \times U_{\sigma^{(L-1)}} \ . \tag{35}$$

The symbol sequence $\underline{\sigma} = (\sigma^{(0)}, \sigma^{(1)}, \ldots, \sigma^{(L-1)})$ which labels the set (35) takes the spatial degrees of freedom into account. Each set is mapped to I^L by the system $T_{\varepsilon=0}$. To take the temporal aspects of the motion into account we again introduce finer partitions of the phase space in terms of cylinder sets. Since we consider still the uncoupled case the cylinder sets are given by direct products of the single site quantities

$$U_{\underline{\sigma}_0\underline{\sigma}_1\cdots\underline{\sigma}_{n-1}} = \left\{\underline{x} \in I^L \,|\, T^k_{\varepsilon=0}(\underline{x}) \in U_{\underline{\sigma}_k}, 0 \le k \le n-1\right\} \ . \tag{36}$$

Obviously the cylinder sets are linked with the dynamics through the property (cf. (8))

$$T_{\varepsilon=0}(U_{\underline{\sigma}_0\underline{\sigma}_1\cdots\underline{\sigma}_{n-1}}) = U_{\underline{\sigma}_1\cdots\underline{\sigma}_{n-1}} \tag{37}$$

These geometric features are summarised in Fig. 6. In particular, the product structure of the cylinder sets reflects the fact that our dynamical system possesses no spatial coupling. Cylinder sets, i.e. roughly speaking the space time orbits, are labelled by two-dimensional spin lattices $\underline{\sigma}_0\underline{\sigma}_1\ldots\underline{\sigma}_{n-1}$. One dimension of the spin lattice corresponds to the temporal evolution whereas the second dimension takes the spatial extension of the dynamical system into account.

The invariant measure (15) of the single site map carries over to our dynamical system as well. Using the condition (21) we have

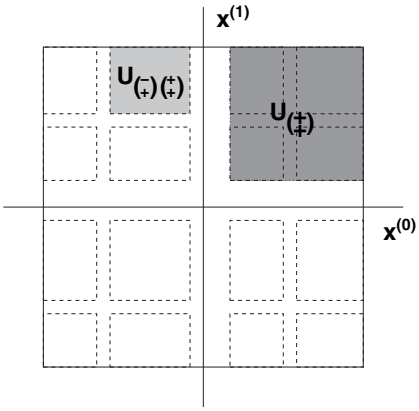


Fig. 6. Diagrammatic view of the first generation of cylinder sets (*dashed rectangles*) of the uncoupled map lattice (34) for $L = 2$. A cylinder set $U_{\underline{\sigma}\,\underline{\tilde{\sigma}}}$ and its image $U_{\underline{\tilde{\sigma}}}$ are *grey shaded*

$$\mu_{\varepsilon=0}(U_{\underline{\sigma}_0\underline{\sigma}_1\cdots\underline{\sigma}_{n-1}}) = h_{\underline{\sigma}_0\underline{\sigma}_1}\frac{p_{\underline{\sigma}_0\underline{\sigma}_1}}{\alpha}\frac{p_{\underline{\sigma}_1\underline{\sigma}_2}}{\alpha}\cdots\frac{p_{\underline{\sigma}_{n-3}\underline{\sigma}_{n-2}}}{\alpha}\,\nu_{\underline{\sigma}_{n-2}\underline{\sigma}_{n-1}} \tag{38}$$

where the weights are defined in terms of the Jacobian $DT_{\varepsilon=0}$ of the uncoupled map

$$p_{\underline{\sigma}\,\underline{\tilde{\sigma}}} = \prod_{\nu=0}^{L-1}\frac{1}{\gamma_{\sigma^{(\nu)}\tilde{\sigma}^{(\nu)}}} = \left|\det\left(DT_{\varepsilon=0}(\underline{x})|_{\underline{x}\in U_{\underline{\sigma}\,\underline{\tilde{\sigma}}}}\right)\right|^{-1}\ . \tag{39}$$

The remaining quantities entering the invariant measure (38) are likewise given as products of the single site quantities and are thus determined by the associated eigenvalue problem[11] (cf. (16) and (17))

$$\alpha h_{\underline{\sigma}_1\underline{\sigma}_2} = \sum_{\underline{\sigma}_0} h_{\underline{\sigma}_0\underline{\sigma}_1}p_{\underline{\sigma}_0\underline{\sigma}_1} \tag{40a}$$

$$\alpha\nu_{\underline{\sigma}_0\underline{\sigma}_1} = \sum_{\underline{\sigma}_2} p_{\underline{\sigma}_0\underline{\sigma}_1}\nu_{\underline{\sigma}_1\underline{\sigma}_2}\ . \tag{40b}$$

Following the spirit of the previous section we can rewrite the measure (38) in terms of a canonical weight of a two-dimensional spin Hamiltonian (cf. (28))

$$\begin{aligned} H_{\underline{\sigma}_0\underline{\sigma}_1\cdots\underline{\sigma}_{n-1}} &= -\ln\mu_{\varepsilon=0}(U_{\underline{\sigma}_0\underline{\sigma}_1\cdots\underline{\sigma}_{n-1}}) \\ &= -\ln h_{\underline{\sigma}_0\underline{\sigma}_1} - \ln\nu_{\underline{\sigma}_{n-2}\underline{\sigma}_{n-1}} \\ &\quad + \sum_{k=0}^{n-2}\sum_{\nu=0}^{L-1}\ln\gamma_{\sigma_k^{(\nu)}\sigma_{k+1}^{(\nu)}} + (n-2)L\ln\alpha\ . \end{aligned} \tag{41}$$

[11] In general, the largest eigenvalue scales exponentially with the system size, i.e. $\sqrt[L]{\alpha}$ tends to a finite value in the thermodynamic limit.

Thus we obtain the statistical mechanics of uncoupled spin chains. The particular boundary conditions are again determined by the eigenvectors of the transfer matrix. No phase transition will occur in this model. But we suspect that the corresponding two-dimensional spin Hamiltonian may exhibit equilibrium phase transitions when a spatial coupling is introduced in the coupled map lattice. Hence the dynamical system may develop different ergodic components in the thermodynamic limit.

In order to introduce a nearest neighbour coupling, while keeping the symbolic dynamics as simple as possible, we consider a piecewise linear coupled map lattice where the local slopes depend on the neighbouring symbols. A simple model is obtained when an unidirectional coupling with periodic boundary conditions $x^{(L)} = x^{(0)}$ is considered. Using the abbreviations (cf. (28))

$$\Gamma_{\sigma^{(\nu)}\tilde{\sigma}^{(\nu)}\sigma^{(\nu+1)}} = \exp\left(-h\sigma^{(\nu)} - J\sigma^{(\nu)}\left(\tilde{\sigma}^{(\nu)} + \sigma^{(\nu+1)}\right) + e_0\right) \tag{42}$$

and (cf. (4))

$$G_{-\sigma}(x) = \begin{cases} \Gamma_{--\sigma}(x+c) & \text{if} \quad x < -c \\ \Gamma_{-+\sigma}(x+c) & \text{if} \quad -c < x < 0 \end{cases} \tag{43a}$$

$$G_{+\sigma}(x) = \begin{cases} \Gamma_{+-\sigma}(x-c) & \text{if} \quad 0 < x < c \\ \Gamma_{++\sigma}(x-c) & \text{if} \quad c < x \end{cases} \tag{43b}$$

we define the coupled map lattice by[12]

$$(T_\varepsilon(\underline{x}))^{(\nu)} = \begin{cases} G_{--}(x^{(\nu)}) & \text{if} \quad x^{(\nu)} < 0 \quad \text{and } x^{(\nu+1)} < 0 \\ G_{-+}(x^{(\nu)}) & \text{if} \quad x^{(\nu)} < 0 \quad \text{and } x^{(\nu+1)} > 0 \\ G_{+-}(x^{(\nu)}) & \text{if} \quad x^{(\nu)} > 0 \quad \text{and } x^{(\nu+1)} < 0 \\ G_{++}(x^{(\nu)}) & \text{if} \quad x^{(\nu)} > 0 \quad \text{and } x^{(\nu+1)} > 0 \end{cases} . \tag{44}$$

Equation (44) is essentially the single site map (3) with the local slope depending on the sign of the right neighbour coordinate. The interaction is mediated by the parameter J (cf. (42)) but we still keep the symbol $T_{\varepsilon=J}$ for the coupled map lattice in order to be consistent with the notation introduced in Sect. 1. The partition on which the coupled map lattice is defined is now given by (cf. (35))

$$V_{\underline{\sigma}} = G^{-1}_{\sigma^{(0)}\sigma^{(1)}}(I) \times G^{-1}_{\sigma^{(1)}\sigma^{(2)}}(I) \times \cdots \times G^{-1}_{\sigma^{(L-1)}\sigma^{(0)}}(I) \tag{45}$$

and the corresponding cylinder sets read (cf. (36))

$$V_{\underline{\sigma}_0\underline{\sigma}_1\cdots\underline{\sigma}_{n-1}} = \left\{\underline{x} \in I^L \,|\, T^k_\varepsilon(\underline{x}) \in V_{\underline{\sigma}_k}, 0 \le k \le n-1\right\} . \tag{46}$$

To distinguish between the quantities of the uncoupled and the coupled map lattice we use the symbol V for the sets of the latter model. Of course the

[12] It is still possible to cast the coupled map lattice in the form (1b). But we refrain from writing down the coupling function explicitly.

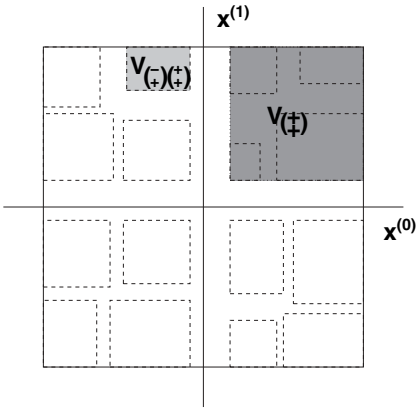


Fig. 7. Diagrammatic view of the first generation of cylinder sets (*dashed rectangles*) of the coupled map lattice (44) for $L = 2$. A cylinder set $V_{\underline{\sigma}\,\underline{\tilde{\sigma}}}$ and its image $V_{\underline{\tilde{\sigma}}}$ are *grey shaded*. Cf. Fig. 6 for the case without coupling

dynamical constraint (37) is again valid if the quantities of the coupled map lattice are considered. A diagrammatic view of the cylinder sets is shown in Fig. 7. Due to the finite coupling the product structure of the cylinder sets is removed.

Since the map lattice T_ε is piecewise linear, the Jacobian yields

$$\begin{aligned} &\left|\det(DT_\varepsilon(\underline{x}))|_{\underline{x}\in V_{\underline{\sigma}\,\underline{\tilde{\sigma}}}}\right| \\ &= \prod_{\nu=0}^{L-1} \Gamma_{\sigma^{(\nu)}\tilde{\sigma}^{(\nu)}\sigma^{(\nu+1)}} \\ &= \exp\left(-h\sum_{\nu=0}^{L-1}\sigma^{(\nu)} - J\sum_{\nu=0}^{L-1}\sigma^{(\nu)}(\tilde{\sigma}^{(\nu)}+\sigma^{(\nu+1)}) + Le_0\right) . \end{aligned} \tag{47}$$

The invariant measure of our model is given by (cf. (38))

$$\mu_\varepsilon(V_{\underline{\sigma}_0\underline{\sigma}_1\cdots\underline{\sigma}_{n-1}}) = h_{\underline{\sigma}_0\underline{\sigma}_1}\frac{p_{\underline{\sigma}_0\underline{\sigma}_1}}{\alpha}\frac{p_{\underline{\sigma}_1\underline{\sigma}_2}}{\alpha}\cdots\frac{p_{\underline{\sigma}_{n-3}\underline{\sigma}_{n-2}}}{\alpha}\nu_{\underline{\sigma}_{n-2}\underline{\sigma}_{n-1}} \tag{48}$$

where the weight is determined by the inverse of the local expansion rate (cf. (21)), i.e.

$$p_{\underline{\sigma}\,\underline{\tilde{\sigma}}} = \frac{1}{\left|\det(DT_\varepsilon(\underline{x})|_{\underline{x}\in V_{\underline{\sigma}\,\underline{\tilde{\sigma}}}})\right|} \tag{49}$$

and the corresponding eigenvalue problem (40) determines the remaining parameters of the measure (48). By construction the corresponding Hamiltonian of the two-dimensional symbol lattice is now given by

$$
\begin{aligned}
&H_{\underline{\sigma}_0\underline{\sigma}_1\cdots\underline{\sigma}_{n-1}} \\
&\quad = -\ln \mu_\varepsilon(V_{\underline{\sigma}_0\underline{\sigma}_1\cdots\underline{\sigma}_{n-1}}) \\
&\quad \simeq \sum_{k=0}^{n-1}\sum_{\nu=0}^{L-1}\left(-h\sigma_k^{(\nu)} - J\sigma_k^{(\nu)}\left(\sigma_{k+1}^{(\nu)} + \sigma_k^{(\nu+1)}\right)\right) \\
&\quad + nL(e_0 + \ln \sqrt[L]{\alpha}) \,. \qquad (50)
\end{aligned}
$$

Thus we end up with a nearest neighbour coupled two-dimensional Ising model. For $h = 0$ this model displays a second order phase transition at $J_c = \mathrm{Artanh}(\sqrt{2}-1)$ in the thermodynamic limit. The equilibrium correlations become long ranged, a macroscopic magnetisation appears, and the canonical distribution develops two different ergodic components. These properties are shared by the underlying coupled map lattice. Hence the spatio–temporal correlations do not decay exponentially (cf. (26)) and the time average of $\sum_{\nu=0}^{L-1}\sigma^{(\nu)}/L$ remains finite in the thermodynamic limit (cf. (25)) despite the fact that any finite system coupled map lattice still has a mixing dynamics. The two different ergodic components which develop at the phase transition point just correspond to the pure phases of positive and negative magnetisation, i.e. to trajectories which on average tend to localise either in the left-hand or the right-hand part of the interval I. The ground state of the Hamiltonian (50), i.e. the zero temperature configuration, is given by the fully aligned spin lattice (for $J > 0$). These spin lattices correspond to the two spatially homogeneous fixed points of the map lattice. Thus the phase transition is a macroscopic localisation of the measure at the linearly unstable fixed points. In this respect the coupled map lattice shares some common features with the dynamics of the single site map (cf. Sect. 1). But for the coupled map lattice it is remarkable that the involved fixed points are not marginally stable at the phase transition point and throughout the low-temperature phase. A modulation of the local slope by a factor of four is already sufficient to induce the phase transition.

In principle it is even possible to analyse our model by direct numerical simulations despite we are dealing with repelling sets. As already indicated in Sect. 1 the motion on repelling sets can be generated by an iterated function scheme which is determined by the inverse branches of the coupled map lattice determine. The transition probabilities to generate a state $\underline{\sigma}_n$ provided the history $\underline{\sigma}_k$, $k \geq n+1$ is given read (cf. (31))

$$
\begin{aligned}
w(\underline{\sigma}_n|\underline{\sigma}_{n+1},\underline{\sigma}_{n+2},\ldots) &= \lim_{k\to\infty}\frac{\mu_\varepsilon(V_{\underline{\sigma}_n\underline{\sigma}_{n+1}\cdots\underline{\sigma}_k})}{\mu_\varepsilon(V_{\underline{\sigma}_{n+1}\cdots\underline{\sigma}_k})} \\
&= \frac{h_{\underline{\sigma}_n\underline{\sigma}_{n+1}}}{h_{\underline{\sigma}_{n+1}\underline{\sigma}_{n+2}}}\frac{p_{\underline{\sigma}_n\underline{\sigma}_{n+1}}}{\alpha} \qquad (51)
\end{aligned}
$$

where the representation (48) of the measure has been used. Since the explicit form of the eigenvector of the transfer matrix is involved the corresponding

stochastic spin dynamics is already fairly complicated. Thus we do not pursue this approach here.

Our model system shows that based on symbolic dynamics it is indeed possible to identify a qualitative change in the dynamics which appears only in the limit of infinite system size. By construction the type of phase transition described here is contained in the Ising universality class. For different choices of the single site map and the local slopes (42) quite general Hamiltonians can be realised. Thus different types of equilibrium phase transitions may appear in these models. One should however keep in mind that the approach requires to some extent a simple symbolic dynamics, since the influence of a complicated grammar on the phase transition behaviour is still unexplored.

3 Coupled Map Lattices and Kinetic Ising Models

The concept described in the previous section mainly focuses on stationary properties of dynamical systems like invariant measures or (stationary) correlation functions. Transient dynamics like relaxation towards equilibrium is difficult to tackle within such an approach. In contrast, coarse grained descriptions of the dynamics in terms of suitable partitions of phase space that keep the time as a dynamical variable are useful to cope with such problems. We will sketch that such ideas are capable to deal with the dynamics of coupled map lattices from a different perspective.

For the purpose of illustration let us consider a spatially one-dimensional coupled map lattice with single site maps motivated by the Miller–Huse model, cf. Sect. 1. Hence we take the following double tent map defined on $I = [-1, 1]$ (cf. Fig. 8)

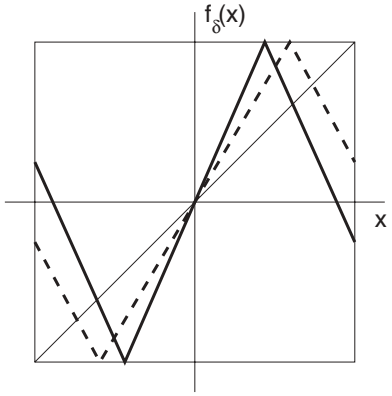


Fig. 8. Single site map, (52), for two different values of the parameter δ. *Solid line*: $\delta < 0$, *broken line* $\delta > 0$

$$f_\delta(x) = \begin{cases} -2 - x/a & \text{if} \quad -1 \le x \le -a \\ x/a & \text{if} \quad -a < x < a \\ 2 - x/a & \text{if} \quad a \le x \le 1 \end{cases} \tag{52}$$

where $1/a = 2 - \delta$ denotes the modulus of the slope. We introduce a small parameter $|\delta| \ll 1$ which will be useful later on for applying some analytical expansion. For $\delta = 0$ the two subintervals $I_+ = [0, 1]$ and $I_- = [-1, 0]$ are invariant sets. The single site map induces a hopping between these two intervals if $\delta < 0$ while hopping is suppressed for $\delta > 0$. On the level of a coarse grained description in terms of symbols $\sigma_n = \mathrm{sign}(x_n)$ the motion may be considered as a kinetic single spin model. The transition emerging at $\delta = 0$ is a plain bifurcation, a so called crises [46], and the motion for $\delta < 0$ is usually called crisis induced intermittency. Nothing like a proper phase transition takes place at $\delta = 0$.

Placing these maps on a chain of length L, introducing the nearest neighbour coupling (1a), and choosing periodic boundary conditions we obtain a kind of minimal model [21]

$$x_{n+1}^{(\nu)} = (T_{\varepsilon,\delta}(\underline{x}_n))^{(\nu)} = (1-\varepsilon) f_\delta(x_n^{(\nu)}) + \frac{\varepsilon}{2}\left[f_\delta(x_n^{(\nu+1)}) + f_\delta(x_n^{(\nu-1)})\right] \ . \tag{53}$$

It is our goal to analyse the dynamics of (53) using perturbation expansions for small $|\delta|$ and ε. Thus let us first comment on the trivial case $\delta = \varepsilon = 0$. Then each cube

$$I_{\underline{\sigma}} = I_{\sigma^{(0)}} \times I_{\sigma^{(1)}} \times \cdots \times I_{\sigma^{(L-1)}} \tag{54}$$

labelled by a symbol string $\underline{\sigma} = (\sigma^{(0)}, \ldots, \sigma^{(L-1)}) \in \{-, +\}^L$ is an invariant set with respect to the uncoupled map lattice $T_{\varepsilon=0,\delta=0}$. The dynamics in each cube is mixing. If we turn on the coupling and the deformation parameter δ transitions between the cubes may be induced. For instance considering an ensemble of initial conditions it seems tempting to describe the motion on a coarse grained level by the probability $p_n(\underline{\sigma})$ that at time n the trajectory visits cube $I_{\underline{\sigma}}$. Based on quite general arguments one may conclude that these probabilities obey a master equation

$$p_{n+1}(\underline{\sigma}) = p_n(\underline{\sigma}) + \sum_{\tilde{\underline{\sigma}}} \left[w(\tilde{\underline{\sigma}} \to \underline{\sigma}) p_n(\tilde{\underline{\sigma}}) - w(\underline{\sigma} \to \tilde{\underline{\sigma}}) p_n(\underline{\sigma})\right] \ . \tag{55}$$

Formal derivations of such an equation from first principles can be obtained e.g. by applying standard projection operator techniques borrowed from non-equilibrium statistical physics [47, 48]. As for the structure of the master (55) we just recall that the two contributions in the sum just yield the gain and the loss due to transitions. The transition $\underline{\sigma} \to \underline{\sigma}$ drops from our considerations. For large n the solution of (55) tends towards a stationary state $p_*(\underline{\sigma})$ which in fact may be identified with $\mu(I_{\underline{\sigma}})$ when μ denotes the SRB measure of the coupled map lattice. The essential approximation in writing down (55) consists in neglecting non-Markovian contributions. Since in the perturbative

regime $|\delta|, \varepsilon \ll 1$ transitions between different cubes are rare, the mixing dynamics in each cube ensures that the memory between different transitions is destroyed and the Markovian approach in terms of (55) becomes feasible. The essential step remains to be the evaluation of the transition probabilities $w(\underline{\sigma} \to \underline{\tilde{\sigma}})$ quantifying transitions from cube $I_{\underline{\sigma}}$ to cube $I_{\underline{\tilde{\sigma}}}$.

Let us emphasise the difference between the present concept and the approach used previously in Sect. 2. The description in terms of the master (55) relies on a spatial coarse graining of the dynamics. We do not require that the partition in terms of the cubes (54) fulfills some Markov property. Hence we do not resolve finer scales in phase space in terms of cylinder sets (cf. (46)). Thus the probabilities $p_n(\underline{\sigma})$ are not capable to resolve the statistics within a cube. But the approach via (55) still contains the time as a dynamical variable and thus can deal explicitly with transient dynamics and the relaxation towards equilibrium.

We are now going to estimate the transition probability for a fixed transition $\underline{\sigma} \to \underline{\tilde{\sigma}}$ in the limit of small coupling and deformation δ. Actually we will introduce a suitable notion for a transition so that finally the master equation is valid. The obvious definition $\underline{x}_n \in I_{\underline{\sigma}}$, $\underline{x}_{n+1} \in I_{\underline{\tilde{\sigma}}}$ fails since it does not refer to the time scale separation which accounts for the validity of the Markov approximation. In the perturbative regime transitions are rare as the trajectory spends most time in the interior of a cube. Let us assume that a particular coordinate $x^{(\nu)}$ at a fixed lattice site ν is going to change sign, i.e. it is going to induce the transition. Then depending on $\sigma^{(\nu)} = \mathrm{sign}(x^{(\nu)})$ the coordinate is either close to the maximum or to the minimum of the single site map

$$x^{(\nu)} = a\sigma^{(\nu)} + \mathcal{O}(\varepsilon, \delta) \tag{56}$$

since f_δ maps such points to the boundary of the interval $I_{\sigma^{(\nu)}}$ (cf. Fig. 8). Let us compute successive iterates up to first order in the small parameters ε and δ. Applying (53) and taking into account that to first order $f_\delta(x^{(\nu)}) = (1 - 2|x^{(\nu)} - a\sigma^{(\nu)}|)\sigma^{(\nu)}$ we obtain for the first iterate at the site ν and its two neighbours

$$\begin{aligned}\left(T_{\varepsilon,\delta}(\underline{x})\right)^{(\nu)} &= \sigma^{(\nu)} - \sigma^{(\nu)}\left(\varepsilon + 2|x^{(\nu)} - a\sigma^{(\nu)}|\right) \\ &\quad + \frac{\varepsilon}{2}\sum_{\rho=\pm 1} f_0(x^{(\nu+\rho)}) + \mathcal{O}(\varepsilon^2, \varepsilon\delta, \delta^2)\end{aligned} \tag{57}$$

and

$$\left(T_{\varepsilon,\delta}(\underline{x})\right)^{(\nu\pm 1)} = f_0(x^{(\nu\pm 1)}) + \mathcal{O}(\varepsilon, \delta)\ . \tag{58}$$

Observing that the coordinate (57) is close to the boundary a similar computation yields for the second iterate

$$\begin{aligned}\left(T^2_{\varepsilon,\delta}(\underline{x})\right)^{(\nu)} &= \sigma^{(\nu)}\left(\delta + 2\varepsilon + 4|x^{(\nu)} - a\sigma^{(\nu)}|\right) \\ &\quad - \varepsilon\sum_{\rho=\pm 1}\left[f_0(x^{(\nu+\rho)}) - \frac{1}{2}f_0^2(x^{(\nu+\rho)})\right] + \mathcal{O}(\varepsilon^2, \varepsilon\delta, \delta^2)\end{aligned} \tag{59}$$

and

$$\left(T^2_{\varepsilon,\delta}(\underline{x})\right)^{(\nu\pm1)} = f_0^2(x^{(\nu\pm1)}) + \mathcal{O}(\varepsilon,\delta)\ . \tag{60}$$

Proceeding in the same way we obtain for the ℓth iterate

$$\begin{aligned}\left(T^\ell_{\varepsilon,\delta}(\underline{x})\right)^{(\nu)} = 2^{\ell-2}\sigma^{(\nu)}\Big(\delta + 2\varepsilon + 4|x^{(\nu)} - a\sigma^{(\nu)}| \\ - \varepsilon\sigma^{(\nu)} \sum_{\rho=\pm1}\Big[f_0(x^{(\nu+\rho)}) - 2t_\ell(x^{(\nu+\lambda)})\Big]\Big) + \mathcal{O}(\varepsilon^2,\varepsilon\delta,\delta^2)\end{aligned} \tag{61}$$

where

$$t_\ell(x) = \sum_{k=2}^{\ell} \frac{1}{2^k} f_0^k(x)\ . \tag{62}$$

Equation (61) describes the asymptotic growth of the coordinate $x^{(\nu)}$ as long as its value stays in $[-a,a]$. From the definition (62) it is clear that the function t_ℓ approaches rapidly the limit

$$t_\infty(x) = \lim_{\ell\to\infty} t_\ell(x) \tag{63}$$

which is frequently called the Takagi function (cf. Fig. 9). Hence using this asymptotic property the coordinate $x^{(\nu)}$ changes sign from $\sigma^{(\nu)}$ to $-\sigma^{(\nu)}$ if

$$\delta + 2\varepsilon + 4|x^{(\nu)} - a\sigma^{(\nu)}| - \varepsilon\sigma^{(\nu)} \sum_{\rho=\pm1}\Big[f_0(x^{(\nu+\rho)}) - 2t_\infty(x^{(\nu+\rho)})\Big] < 0\ . \tag{64}$$

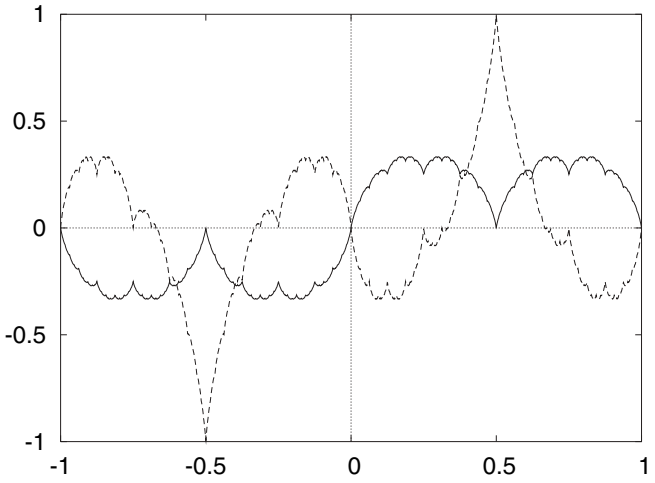


Fig. 9. *Solid line*: Takagi function $t_\infty(x)$, (63). *Broken line*: $f_0(x) - 2t_\infty(x)$ (cf. (66) and (68))

Thus the condition for the occurrence of a transition just depends on the neighbouring coordinates[13], i.e. on the nearest neighbour symbols $\sigma^{(\nu\pm1)}$. According to this observation the transitions can be grouped into three distinct types

$$\text{type I:} \quad +\,{+}{+} \to +\,{-}\,{+} \qquad -\,{-}\,{-} \to -\,{+}\,{-} \tag{65a}$$

$$\text{type II:} \quad -\,{+}{+} \to -\,{-}\,{+} \qquad +\,{-}\,{-} \to +\,{+}\,{-} \tag{65b}$$

$$\text{type III:} \quad +\,{-}{+} \to +\,{+}\,{+} \qquad -\,{+}\,{-} \to -\,{-}\,{-} \tag{65c}$$

Only those transitions $\underline{\sigma} \to \tilde{\underline{\sigma}}$ enter the master (55) where a single symbol of the chain $\underline{\sigma}$ is flipped. Master equations with such special transition probabilities are often called kinetic Ising models. The dynamics of such an equation can be generated in terms of a stochastic spin dynamics according to the (local) transition rates of the master equation [49, 50]. Therefore even numerical simulations become feasible. They are widely used in the context of solid state physics.

The condition (64) determines which type of transition (65) is really possible and enters the sum in (55) with finite probability. Obviously the optimal choice to fulfil the condition (64) is given by $x^{(\nu)} = a\sigma^{(\nu)}$ indicating that the transition starts in the centre of the interval. Optimal choices for the nearest neighbour coordinates have to maximise the expression containing the Takagi function, subjected to the constraint which type of transition is considered, i.e. which values for the nearest neighbour symbols $\sigma^{(\nu\pm1)} = \mathrm{sign}(x^{(\nu\pm1)})$ are assumed. Introducing the abbreviations

$$b_+ = \sup\{f_0(x) - 2t_\infty(x) \,|\, x \in [0,1]\} \tag{66a}$$

$$b_- = \sup\{f_0(x) - 2t_\infty(x) \,|\, x \in [-1,0]\} \tag{66b}$$

the condition (64) yields the constraints

$$\text{type I:} \quad \delta < -2\varepsilon + 2\varepsilon b_+ \tag{67a}$$

$$\text{type II:} \quad \delta < -2\varepsilon + \varepsilon(b_+ + b_-) \tag{67b}$$

$$\text{type III:} \quad \delta < -2\varepsilon + 2\varepsilon b_- \tag{67c}$$

The numerical values of the constants (66) are computed straightforwardly by employing the self similar features of the Takagi function (cf. Fig. 9)

$$b_+ = 1, \qquad b_- = 1/3\,. \tag{68}$$

Hence the conditions (67) split the ε–δ parameter plane into four regions where different types of transitions occur (cf. Fig. 10 and [21] for a detailed analysis of the underlying kinetic Ising model):

[13] In higher orders of the perturbation expansion other lattice sites will enter as well.

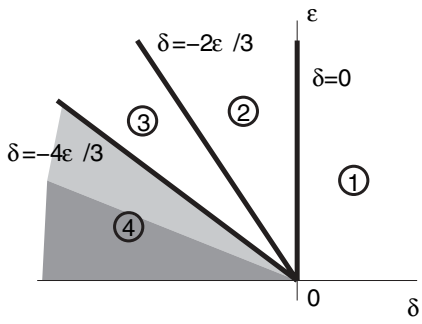


Fig. 10. Diagrammatic view of the bifurcation diagram of the coupled map lattice (53) according to the conditions (67). Labels refer to the regions mentioned in the text. *Greyshading* indicates the type of coupling according to (72), antiferromagnetic (*light*), ferromagnetic (*dark*)

- Region 1 ($\delta > 0$): According to (67) and (68) no transition is possible. No spin flips occur, i.e. all transition probabilities of the master (55) vanish. Each cube $I_{\underline{\sigma}}$ contains an attractor of the coupled map lattice.
- Region 2 ($-2\varepsilon/3 < \delta < 0$): Only transitions of type I are permitted. Those cubes $I_{\underline{\sigma}}$ contain an attractor where the symbol chain $\underline{\sigma}$ does not contain three successive symbols of the same type.
- Region 3 ($-4\varepsilon/3 < \delta < -2\varepsilon/3$): Transitions of type I and II are permitted. The dynamics settles on an alternating symbol sequence $\underline{\sigma} = (+,-,+,-,\ldots)$ or $\underline{\sigma} = (-,+,-,+,\ldots)$, if we consider for simplicity systems of even system size L. Thus there are two attractors
- Region 4 ($\delta < -4\varepsilon/3$): All types of transitions are feasible and the attractor encompasses all cubes.

Roughly speaking, decreasing δ increases the number of possible transitions and decreases the number of attractors of the coupled map lattice. One should however keep in mind that these changes are just plain bifurcations which change the topology in the phase space. They are not linked to (proper) phase transitions as the size of the system does not play an essential role.

While we have so far discussed which transition is possible we may in addition estimate the numerical value of the corresponding transition probability. Condition (64) determines those points in the phase space which trigger a particular transition

$$E_{\sigma^{(\nu-1)},\sigma^{(\nu)},\sigma^{(\nu+1)}} = \Big\{ (x^{(\nu-1)}, x^{(\nu)}, x^{(\nu+1)}) \in I_{\sigma^{(\nu-1)}} \times I_{\sigma^{(\nu)}} \times I_{\sigma^{(\nu+1)}} \,|\, \delta + 2\varepsilon + 4|x^{(\nu)} - a\sigma^{(\nu)}| - \varepsilon\sigma^{(\nu)} \sum_{\rho=\pm 1} \Big[f_0(x^{(\nu+\rho)}) - 2t_\infty(x^{(\nu+\rho)}) \Big] < 0 \Big\} \,. \tag{69}$$

Assuming that due to the mixing properties of the tent map and the small transitions rates the distribution of phase space points stays uniform within each cube we may estimate the transition rate by the (normalised) volume of the transition region (69). The numerical values of the transition rates determine the dynamics of the corresponding kinetic Ising model. Here we focus on the behaviour in region 4 and refer the interested reader for the discussion of the other cases to the literature [21]. In region 4 all three types of transitions (67) are possible with finite transition rates w_I, w_{II} and w_{III}. Actually we have obtained a kinetic Ising model with nearest neighbour interaction (cf. [49]). For the stationary distribution the sum in the master equation (55) must vanish. In our case the even stronger condition of detailed balance holds meaning that each term of the sum vanishes individually[14]

$$w(\tilde{\underline{\sigma}} \to \underline{\sigma}) p_*(\tilde{\underline{\sigma}}) - w(\underline{\sigma} \to \tilde{\underline{\sigma}}) p_*(\underline{\sigma}) = 0 \; . \tag{70}$$

Observing that transitions of type II do not increase the number of $+/-$ pairs in the symbol chain, and that transitions of type I and III are inverse to each other one easily recognises that the stationary solution can be written as

$$p_*(\underline{\sigma}) = \frac{1}{Z} \exp\left(J \sum_{\nu=0}^{L-1} \sigma^{(\nu)} \sigma^{(\nu+1)} \right) \tag{71}$$

when we define the effective exchange constant by the ratio of the transition probabilities

$$J = \frac{1}{4} \ln\left(\frac{w_{III}}{w_I} \right) \quad . \tag{72}$$

In fact, the validity of (70), (71), and (72) is easily checked a posteriori. Close to the boundary of region 3 the transition rate w_{III} is small and the coupling (72) is predominantly antiferromagnetic, in accordance with the stationary state in region 3. Within region 4 w_{III} may increase and the coupling may turn out to become ferromagnetic (cf. Fig. 10). Above all the stationary state in region 4 is a finite temperature canonical ensemble of an Ising chain. Of course no phase transition will occur in this one-dimensional setup.

Since the one-dimensional coupled map lattice (53) can be described in terms of a one-dimensional kinetic Ising model it seems tempting to investigate (proper) phase transitions by considering the spatially two-dimensional case. But here one has to add an important reservation. The analysis of the model (53) resulted in a coarse grained description which obeys detailed balance (70). Such a condition requires constraints on the transition rates which crucially depend on the underlying coupled map lattice. Taking for instance a one-dimensional coupled map lattice with asymmetric coupling, one obtains transition rates violating detailed balance (cf. [52] for the analysis of the

[14] Actually this condition poses a constraint on the transition probabilities in terms of a potential condition [51].

unidirectionally coupled case). Thus in general one may end up with nonequilibrium kinetic Ising models. In such cases phase transitions are possible even in the spatially one-dimensional setup [53]. A similar phenomenon occurs for the obvious generalisation of the model (53) to two spatial dimensions. Again one obtains a master equation violating detailed balance. Thus a proper understanding of phase transitions in coupled map lattices calls for a deeper understanding of nonequilibrium models in statistical mechanics. In fact, as proven in [6] one may end up with a coarse grained description which is more general than a kinetic Ising model. In such cases one has to introduce carefully the concept of a (proper) phase transition since the notions from equilibrium statistical mechanics do not apply directly. Above all more research even for quite simple models is still needed in order to classify phase transitions in coupled map lattices properly.

References

1. M. C. Cross and P. C. Hohenberg, *Pattern–formation outside of equilibrium*, Rev. Mod. Phys. **65**, 851 (1993). 33
2. J. P. Crutchfield and K. Kaneko, in *Directions in Chaos*, edited by H. Bai-Lin (World Scientific Publ., Singapore, 1987), p. 272. 34
3. K. Kaneko, *Pattern dynamics in spatiotemporal chaos*, Physica D **34**, 1 (1989). 34, 38
4. J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Springer, New York, 1986). 34
5. H. Haken, *Synergetics* (Springer, Berlin, 1977). 34
6. G. Gielis and R. S. MacKay, *Coupled map lattices with phase transition*, Nonlin. **13**, 867 (2000). 34, 35, 40, 62
7. K. Kaneko, *Theory and applications of coupled map lattices* (Wiley, Chichester, 1993). 35
8. L. A. Bunimovich and Y. G. Sinai, *Spacetime chaos in coupled map lattices*, Nonlin. **1**, 491 (1988). 35, 39
9. J. Bricmont and A. Kupiainen, *Coupled analytic maps*, Nonlin. **8**, 379 (1995). 35, 39
10. C. Maes and A. V. Moffaert, *Stochastic stability of weakly coupled lattice maps*, Nonlin. **10**, 715 (1997). 35, 39
11. T. Fischer and H. H. Rugh, *Transfer operators for coupled analytic maps*, Erg. Theor. Dyn. Syst. **20**, 109 (2000). 35, 39
12. *Collective dynamics of nonlinear and disordered systems*, edited by G. Radons, P. Häussler, and W. Just (Springer, Berlin, 2004). 36
13. K. Kaneko, *Mean field fluctuation of a network of chaotic elements*, Physica D **55**, 368 (1992). 36
14. K. Kaneko, *Globally coupled chaos violates the law of large numbers but not the central–limit theorem*, Phys. Rev. Lett. **65**, 1391 (1990). 36
15. E. Järvenpää, *A SRB–measure for globally coupled circle maps*, Nonlin. **10**, 1435 (1997). 36
16. S. V. Ershov and A. B. Potapov, *On the mean field fluctuations in globally coupled maps*, Physica D **86**, 523 (1995). 36, 40
17. T. Chawanya and S. Morita, *On the bifurcation structure of the mean–field fluctuation in the globally coupled tent map system*, Physica D **116**, 44 (1998). 36

18. G. Keller, *An ergodic theoretic approach to mean field coupled maps*, Prog. Prob. **46**, 183 (2000). 36
19. W. Just, *Globally coupled maps: phase transitions and synchronization*, Physica D **81**, 317 (1995). 36, 37
20. K. Kaneko, *Supertransients, spatiotemporal intermittency and stability of fully developed spatiotemporal chaos*, Phys. Lett. A **149**, 105 (1990). 36
21. F. Schmüser, W. Just, and H. Kantz, *On the relation between coupled map lattices and kinetic Ising models*, Phys. Rev. E **61**, 3675 (2000). 37, 56, 59, 61
22. J. Miller and D. A. Huse, *Macroscopic equilibrium from microscopic irreversibility in a chaotic coupled map lattice*, Phys. Rev. E **48**, 2528 (1993). 38
23. H. Sakaguchi, *Phase transition in coupled Bernoulli maps*, Prog. Theor. Phys. **80**, 7 (1988). 38, 40
24. P. Marcq, H. Chaté, and P. Manneville, *Universality in Ising–like phase transitions of coupled chaotic maps*, Phys. Rev. E **55**, 2606 (1997). 38
25. A. Lemaître and H. Chaté, *Phase ordering and onset of collective behavior in chaotic coupled map lattices*, Phys. Rev. Lett. **82**, 1140 (1999). 38
26. X. J. Wang, *Statistical physics of temporal intermittency*, Phys. Rev. A **40**, 6647 (1989). 39, 47, 49
27. C. S. Hsu, *Cell–to–cell mapping : a method of global analysis for nonlinear systems* (Springer, New York, 1987). 40
28. M. Dellnitz, A. Hohmann, O. Junge, and M. Rumpf, *Exploring invariant sets and invariant measures*, Chaos **7**, 221 (1997). 40
29. M. Dörfle, *Spectrum and eigenfunctions for the Frobenius–Perron operator of the tent map*, J. Stat. Phys. **40**, 93 (1985). 40
30. S. Morita, *Bifurcations in globally coupled chaotic maps*, Phys. Lett. A **211**, 258 (1996). 40
31. C. Beck and F. Schlögl, *Thermodynamics of chaotic systems* (Cambridge University Press, Cambridge, 1995). 40
32. W. Just, in *Collective dynamics of nonlinear and disordered systems*, edited by G. Radons, P. Häussler, and W. Just (Springer, Berlin, 2004). 40
33. K. Falconer, *Techniques in fractal geometry* (Wiley, Chicester, 1997). 40, 47
34. A. Lasota and M. C. Mackey, *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics* (Springer, New York, 1994). 44
35. H. Mori, B. S. So, and T. Ose, *Time-correlation functions of one-dimensional transformations*, Prog. Theor. Phys. **66**, 1266 (1981). 44
36. T. Bohr and D. Rand, *The entropy function for characteristic exponents*, Physica D **25**, 387 (1987). 44
37. G. Pianigiani and J. A. Yorke, *Expanding maps on sets which are almost invariant: Decay and chaos*, Trans. Amer. Math. Soc. **252**, 351 (1979). 44
38. R. Bowen, in *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, Vol. 470 of *Lecture Notes in Mathematics*, edited by A. Dold and B. Eckmann (Springer, Berlin, 1975). 45
39. T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procaccia, and B. I. Shraiman, *Fractal measures and their singularities: The characterization of strange sets*, Phys. Rev. A **33**, 1141 (1986). 46
40. P. Grassberger, R. Badii, and A. Politi, *Scaling laws for invariant measures on hyperbolic and nonhyperbolic attractors*, J. Stat. Phys. **51**, 135 (1988). 46
41. W. Just and H. Fujisaka, *Gibbs Measures and Power Spectra for Type I Intermittent Maps*, Physica D **64**, 98 (1993). 46, 49

42. C. Robinson, *Dynamical systems : stability, symbolic dynamics, and chaos* (CRC Press, Boca Raton, 1995). 49
43. R. Artuso, E. Aurell, and P. Cvitanović, *Recycling of strange sets*, Nonlin. **3**, 325 (1990). 50
44. D. Ruelle, *Thermodynamic formalism* (Addison-Wesley, Reading, 1978). 50
45. W. Just, *Equilibrium phase transitions in coupled map lattices: a pedestrian approach,*, J. Stat. Phys. **105**, 133 (2001). 50
46. C. Grebogi, E. Ott, and J. A. Yorke, *Crisis sudden changes in chaotic attractors and transient chaos*, Physica D **7**, 181 (1983). 56
47. E. Fick and G. Sauermann, *The quantum statistics of dynamic processes* (Springer, Berlin, 1990). 56
48. W. Just, *Projection operator approach to the thermodynamic formalism of dynamical systems*, J. Stat. Phys. **67**, 271 (1992). 56
49. R. J. Glauber, *Time-dependent statistics of the Ising model*, J. Math. Phys. **4**, 294 (1963). 59, 61
50. K. Kawasaki, in *Phase Transitions and Critical Phenomena*, edited by C. Domb (Academic Press, London, 1972), p. 443. 59
51. F. P. Kelly, *Reversibility and Stochastic Networks* (Wiley, Chichester, 1979). 61
52. F. Schmüser and W. Just, *Non-equilibrium behaviour in unidirectionally coupled map lattices*, J. Stat. Phys. **105**, 525 (2001). 61
53. *Nonequilibrium Statistical Mechanics in One Dimension*, edited by V. Privman (Cambridge University Press, Cambridge, 1997). 62

Indecomposable Coupled Map Lattices with Non-unique Phase

R.S. MacKay

Mathematics Institute, University of Warwick, Coventry CV4 7AL, U.K.
mackay@maths.warwick.ac.uk

1 Introduction

A compact topologically mixing uniformly hyperbolic attractor for a diffeomorphism of a finite-dimensional manifold with Hölder continuous derivative carries a unique probability measure describing the long-time statistics of the forward orbits of almost all initial conditions in its basin. This was proved [28] by converting the orbits of the dynamical system into symbol sequences from a finite alphabet (indexed by time) and considering them as states of a statistical mechanical spin chain with a certain interaction energy which decays exponentially with separation, so leading to a unique phase. The results have been extended to many classes of non-uniformly hyperbolic system by different approaches (e.g. [30]).

What about spatially extended dynamical systems? By analogy with statistical mechanics, could increasing from one dimension (time) to more than one (time plus space) permit non-unique probabilistic behaviour (even if the system is indecomposable in an appropriate sense at the topological level)? This question was raised by [13, 5], for example.

Because of the statistical mechanics analogy, I shall refer to the possible probabilistic behaviours of dynamical systems as *phases*. These are probability distributions on the set of all orbits, not just on the state space.

Although infinite systems are probably a fiction, they are a good idealisation to study for such questions, because (i) under conditions to rule out associated spin chains with infinitely many states per site or slowly decaying interaction, finite-dimensional topologically mixing systems have unique phase, and (ii) non-unique phase for an infinite system can be reflected in extraordinarily long transients for large but finite versions (for a proved example, see [29]).

For many researchers the interest in spatially extended dynamical systems focusses on partial differential equations (PDEs), but analysis of PDEs is highly technical, which makes it difficult to address the essential phenomena. The question of non-unique phase is already interesting for spatially discrete

R.S. MacKay: *Indecomposable Coupled Map Lattices with Non-unique Phase*, Lect. Notes Phys. **671**, 65–94 (2005)
www.springerlink.com

systems, however, such as networks of oscillators [19]. I shall address the question in the simplest context: coupled map lattices (CML) [15], where time also is taken to be discrete.

Despite much numerical work suggesting non-unique phase can occur in CML (e.g. [24, 7] and Just's lectures in this school) and some papers claiming to prove it or make it plausible for some examples ([27] comes very close), the paper [10] appears to be the first to give a proved example. I do not count examples where the non-uniqueness arises purely from a bifurcation at the topological level, or where it refers to escape measures on unstable sets (e.g. [14] and Just's lectures).

As has been seen in other lectures in this school, piecewise affine systems are much easier to analyse and yet can exhibit a lot of the phenomenology of smooth ones. Following [10], I shall construct a variety of piecewise affine CML with provably non-unique phase. The examples illustrate not just non-unique invariant measure but also the phenomena of non-trivial collective behaviour, eternal transience and an arrow of time. Although somewhat artificial, it is better to have some concrete examples than nothing at all. An analogy is the 2D Ising model of statistical mechanics, which is a toy model of a ferromagnet but gave the first context in which it was proved that non-unique phase can arise (I don't count mean field approximations).

The examples all have "symbolic dynamics", which again makes them easier to analyse. Systems with symbolic dynamics have a particular class of phase called *Gibbs phases*. I will argue that for systems with symbolic dynamics, Gibbs phases form the physically relevant class, describing the statistics of local observables on the orbits of all "uncertain" initial conditions started a long time in the past.

In more detail, Sect. 2 begins by giving four examples of CML with provably non-unique Gibbs phase. The key idea is to simulate probabilistic cellular automata (PCA) for which analogous non-uniqueness has already been proved. Then I give an appropriate definition of indecomposability and show that these examples are indecomposable. Next, an example is given with an absorbing subset which looks as if it ought to attract the orbit of almost every initial condition, yet there is a large subset of initial conditions which never discover it, despite the system being "pre-indecomposable". Then I show how to build invertible versions of these examples. Finally, I make a time-reversible indecomposable CML with non-reversible phases.

Section 3 motivates the concept of Gibbs phase for CML with symbolic dynamics, by building up from the simpler cases of Markov chains, ordinary dynamical systems and PCAs. Some readers may prefer to read this section before Sect. 2, but I consider it better to go straight to the examples, which have obviously relevant non-unique phases, before going into details of what class of phases should be considered for general CML with symbolic dynamics. The discussion is by no means complete, but I hope it helps.

The notes conclude with some challenges and an appendix on the concept of reversibility for Markov chains and PCA.

2 A Selection of Examples

2.1 CML with Ferromagnetic Phases and Variants

Example 1. Let "space" $S = \mathbb{Z}^2$ and "time" $T = \mathbb{Z}$. For each "site" $s = (s_1, s_2) \in S$, define its neighbourhood $N(s) = \{s, s + (1, 0), s + (0, 1)\}$ (the "North-East-Centre" (NEC) neighbourhood). Take "local state space" $M = [0, 1]$. For "lattice state" $x = (x_s)_{s \in S} \in M^S$ define "lattice symbolic state" $\sigma \in \{+, -\}^S$ by (for each $s \in S$):

$$\sigma_s(x) = \begin{cases} + & \text{if } x_s \geq 1/2 \\ - & \text{if } x_s < 1/2 \end{cases}.$$

For $\varepsilon \in (0, 1]$, apply at each site a piecewise affine map from Fig. 1 according to the symbolic state of the NE neighbours. The slopes are $\frac{2}{\varepsilon}$ and $\frac{2}{2-\varepsilon}$. This defines a CML, even though the influence of neighbours is via only their symbolic state rather than continuously (like the more familiar discrete diffusion). Note that in contrast to common usage, ε represents deviation from deterministic symbolic dynamics rather than coupling strength; the uncoupled case is $\varepsilon = 1$. Given initial condition $x^{-N} \in M^S$ at some large negative time $-N$, denote the state at times $t \in T$ with $t \geq -N$ by x^t.

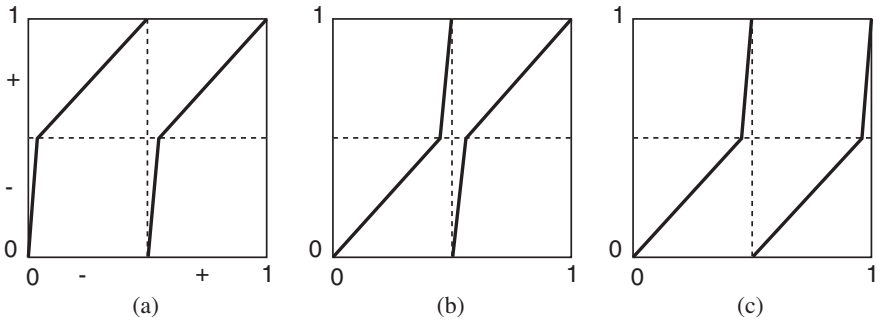


Fig. 1. Local maps for a CML with ferromagnetic phases for $\varepsilon < \varepsilon_c$ (here $\varepsilon = 0.13$), for symbolic state of the NE neighbours: (**a**) $++$, (**b**) $+-$ or $-+$, (**c**) $--$

For ε near 1, this CML has a unique Gibbs phase G. In particular, by the $x \mapsto 1 - x$ symmetry of the CML, $G(\sigma_s(x^t) = +) = \frac{1}{2}$ for all $(s, t) \in S \times T$. In contrast, for ε small enough, there is more than one Gibbs phase. The set of Gibbs phases for a system is convex, so can be described by the *pure* (or "extremal") Gibbs phases, those which can not be written as a non-trivial convex combination of others. Thus I will make the statement in terms of pure Gibbs phases [10]:

There exists $\varepsilon_c \in (0, 1)$ (numerically about 0.18 [3]) such that for all $\varepsilon \in [\varepsilon_c, 1]$ there is a unique Gibbs phase, whereas for each $\varepsilon \in (0, \varepsilon_c)$ there are pure Gibbs phases $G^{\pm}(\varepsilon)$ and $c(\varepsilon) > \frac{1}{2}$ such that for all $(s, t) \in S \times T$,

$$G^+\left(\sigma_s(x^t)=+\right)=G^-\left(\sigma_s(x^t)=-\right)=c\,.$$

Furthermore, $G^\pm$ are space-time translation invariant.

By analogy with statistical mechanics, I say the phases $G^\pm$ are "ferromagnetic". To simplify the following discussion of this example I'll assume that $G^\pm$ are the only pure Gibbs phases (though I'm not aware of a proof).

What is the implication of this result for typical initial conditions? As I'll explain in Sect. 3, for spatially infinite systems I don't know how to interpret "typical", so instead I think in terms of "uncertain" initial conditions. An *uncertain initial condition* is a probability distribution for initial conditions which has absolutely continuous marginal on each finite subset L of S, whose densities ϱ_L satisfy some further conditions. Then the answer is that for any bounded subset Λ of space-time $S\times T$ and initial time sufficiently far in the past, the probability distribution for the components in Λ of the orbits is some convex combination $p_+G^+ + p_-G^-$ with $p_\pm\geq 0$, $p_+ + p_- = 1$, depending on the initial condition, initial time and Λ. In words, you should see samples from the probability distribution G^+ with probability p_+ and from G^- with probability p_-. For most initial conditions drawn from the initial probability distribution, I think this translates into seeing large patches with the statistics of G^+ and large patches with those of G^-, separated by narrow wandering domain walls, though the logical implication is not clear to me.

Example 2. The next example is given by Fig. 2. For $\varepsilon\in(0,\varepsilon_c)$ it has two pure Gibbs states G^{es} (for "even sites") and G^{os} (for "odd sites") with

$$G^{es}\left(\sigma_s(x^t)=(-)^{s_1+s_2}\right)=G^{os}\left(\sigma_s(x^t)=-(-)^{s_1+s_2}\right)=c$$

(the same ε_c and $c(\varepsilon)$ as above). So you see chessboard patterns. The phases are "antiferromagnetic".

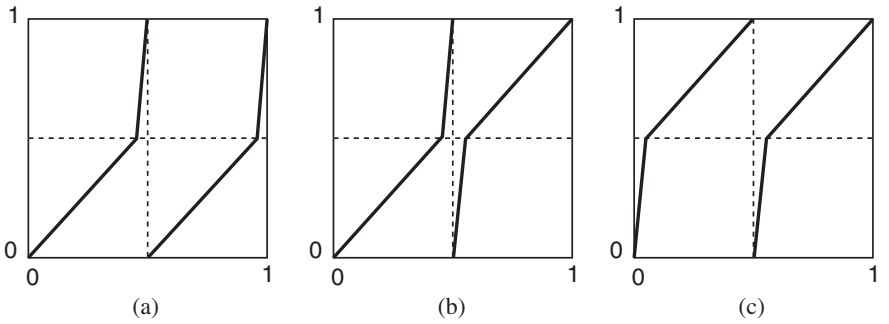


Fig. 2. Local maps for a CML with antiferromagnetic phases for $\varepsilon<\varepsilon_c$, for symbolic state of the NE neighbours: (**a**) $++$, (**b**) $+-$ or $-+$, (**c**) $--$

Example 3. The third example uses the maps in Fig. 3 (equivalently, interchange ε with $2-\varepsilon$ in the first example). For $\varepsilon \in (0, \varepsilon_c)$ it has two extremal Gibbs states G^{et} (for "even time") and G^{ot} (for "odd time") with

$$G^{et}\left(\sigma_s(x^t) = (-)^t\right) = G^{ot}\left(\sigma_s(x^t) = -(-)^t\right) = c\,.$$

So a period-2 behaviour emerges, alternating between a predominance of $+$ and $-$. The phases are "period-2 ferromagnetic". They are examples of "non-trivial collective behaviour" [7, 16] or "asymptotic periodicity" [22]. This is a type of non-uniqueness that would not be apparent if one restricted attention to invariant measures, and yet is highly significant. Hence I think it important to look at measures on the set of orbits, not just invariant ones (on the state space).

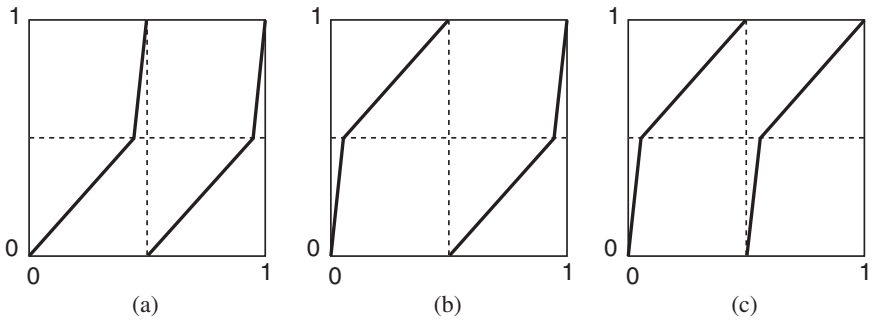


Fig. 3. Local maps for a CML with period-2 ferromagnetic phases for $\varepsilon < \varepsilon_c$, for symbolic state of the NE neighbours: (**a**) $++$, (**b**) $+-$ or $-+$, (**c**) $--$

Example 4. Finally, use the maps in Fig. 4 (equivalently, interchange ε with $2-\varepsilon$ in the second example). For $\varepsilon \in (0, \varepsilon_c)$ it has two extremal Gibbs states G^e and G^o (for "even" and "odd" space-time) with

$$G^e\left(\sigma_s(x^t) = (-)^{s_1+s_2+t}\right) = G^o\left(\sigma_s(x^t) = -(-)^{s_1+s_2+t}\right) = c\,.$$

These phases are "period-2 antiferromagnetic".

2.2 How the Constructions Work

The idea of these constructions is that (i) there are indecomposable probabilistic cellular automata (PCA) with non-unique Gibbs phase, and (ii) any PCA can be simulated by some CML.

A PCA is a stochastic process on an infinite product of finite state spaces generated by transition probabilities given by independent spatially local rules. A good reference on theory and examples of PCA is [29].

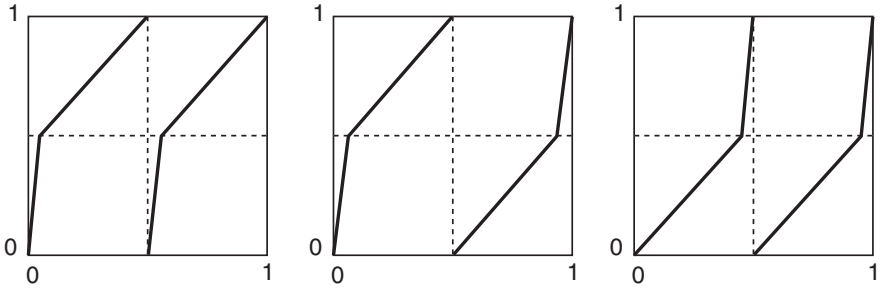


Fig. 4. Local maps for a CML with period-2 antiferromagnetic phases for $\varepsilon < \varepsilon_c$, for symbolic state of the NE neighbours: (**a**) $++$, (**b**) $+-$ or $-+$, (**c**) $--$

The examples of the previous subsection are based on a case of Toom's *majority voter PCA*. Its state space is $\{+,-\}^S$ with $S = \mathbb{Z}^2$. Let m be the majority sign in the NEC neighbourhood of s. Then the σ_s^t are updated independently and synchronously by transition probabilities:

$$P(\sigma_s^{t+1} = m|\ \sigma^t) = 1 - \frac{\varepsilon}{2}$$
$$P(\sigma_s^{t+1} = -m|\ \sigma^t) = \frac{\varepsilon}{2}\ .$$

Toom proved that there is $\varepsilon_c \in (0,1)$ such that starting from the all $+$ state at $t = 0$, for $\varepsilon \in (0, \varepsilon_c)$ there is $c > \frac{1}{2}$ such that for all $s \in S$ and $t \in Z_+$, $P(\sigma_s^t = +) \geq c$ (and converges to c as $t \to \infty$) [29, 20]. I will not go into the proof here, except to remark that it uses a stochastic analogue of monotonicity, a concept that has been seen to be useful in other lectures at this school. Specifically, for all $s \in S$,

$$P(\sigma_s^{t+1} = +|\ \sigma^t = \sigma') \geq P(\sigma_s^{t+1} = +|\ \sigma^t = \sigma'')$$

whenever $\sigma' \geq \sigma''$ in the partial order defined by $\sigma'_s \geq \sigma''_s$ for all $s \in S$ in the order $+ > -$. This property is called "attractivity" in some of the literature.

To simulate a PCA by a CML, if the PCA has N states for each site ($N = 2$ here) and transition probabilities $p_{ij}(n)$ from state i to j at site s given the state n of the neighbours (excluding s itself), use for each n a piecewise affine map of $M = [0,1]$ to itself subdivided into N intervals of equal length to represent the states, with slopes $\frac{1}{p_{ij}(n)}$ from the ith subinterval to the jth. One can allow the lengths of the subintervals to vary, with corresponding modifications of the slopes (e.g. [27]), but ours seems the easiest general prescription (Sakaguchi's produces a second order CML, i.e. x_s^{t+1} depends on x_s^{t-1} as well as x^t, and is less flexible in the set of PCAs it can simulate).

In what sense does the CML simulate the PCA? Firstly, if one considers probability distributions on M^S with constant marginal density on each set corresponding to the symbolic state of a finite subset of S, then the induced

map on the densities preserves this property and is precisely the action of the PCA on probabilities. There is a deeper sense, however. Under iteration of the induced map on probability measures on M^S, for any finite subset $L \subset S$, the marginal density ρ_L of any uncertain initial condition on $[0,1]^S$ becomes asymptotically constant on each cylinder set (i.e. set corresponding to a given symbolic state on L) [10].

Knowledgeable readers might have expected me to use a 2D kinetic Ising model instead of Toom's PCA. This will be done in Sect. 2.7, but Toom's PCA has the advantage of robustness: one can modify the transition probabilities somewhat and still keep the phenomenon of non-unique phase (e.g. [3]). This is in contrast to kinetic Ising models where breaking the $+/-$ symmetry reduces them generically to a unique phase.

Toom's PCA is 2D. Are there 1D examples with non-unique phase (and yet indecomposable in a good sense)? Yes. Indeed there is a 1D PCA with infinitely many pure phases [9], but it is complicated and not entirely explicit (see [12] for an introduction to the paper)! It would be nice to have a simpler 1D example, and two pure phases would suffice.

2.3 Indecomposability

It is clear that non-unique phase can arise in trivial ways if we do not require the system to be indecomposable in some sense. If a dynamical system has more than one attractor there will be at least one phase for each. Even if there is one attractor but it consists of several pieces which are cyclically permuted there will be at least one phase for each of the pieces in which the orbits can start (some people would want to identify them because they give the same time-averages but I think it is useful to distinguish them).

The same goes for Markov chains. There it is usual to require irreducibility and aperiodicity, because any Markov chain can be reduced to pieces with these properties. A Markov chain is *irreducible* if it has no proper forward invariant subset; equivalently, if for all states i, j it is possible to get from i to j in a positive number of steps. An irreducible Markov chain is *aperiodic* if there is N such that for all i, j it is possible to get from i to j in precisely N steps (if not then it is the "composition" of a deterministic periodic cycle with an aperiodic irreducible Markov chain). An aperiodic irreducible finite-state Markov chain has a globally attracting invariant measure, thus a unique phase, whereas a period-M irreducible finite-state Markov chain has a unique invariant measure but M phases.

It is appropriate here to make a warning about terminology. Many probabilists (though not all, e.g. [18]) say a system with a globally attracting invariant measure is *ergodic*, whereas dynamicists say it is *mixing* and use the word "ergodic" for an invariant measure that can not be decomposed into two (i.e. an extremal invariant measure).

For dynamical systems, a good notion of indecomposability is *topological mixing*: for any two non-empty open sets A, B in the state space, the image of A intersects B at all large enough times. This could be applied to CML by considering product topology on M^S.

A stronger notion of indecomposability, however, is Bowen's *specification property* (e.g. [17]): for all $\delta > 0$ there exists $D_0 \in \mathbb{Z}_+$ such that for all finite sequences of orbit segments $x(i), i = 1, \ldots, I$ on consecutive time intervals $[a_i, b_i]$ with separations $a_{i+1} - b_i \geq D_0$ there exists an orbit y agreeing to within δ with $x(i)$ at each time in $[a_i, b_i]$ for each $i \in \{1, \ldots, I\}$.

For CML, an appropriate version of the specification property should refer to space as well as time. Thus I say a CML on a space S is *indecomposable* if for all $\delta > 0$ there exists $D_0 \in \mathbb{R}_+$ such that for all finite sets of bounded subsets $A_i \subset S \times T$, $i = 1, \ldots, I$, with separations at least D_0, and orbits $x(i)$, there exists an orbit y agreeing to within δ with $x(i)$ on A_i for each i.

The *separation* of two subsets $A, B \subset S \times T$ is defined by $D(A, B) = \inf\{d(a, b) : a \in A, b \in B\}$. For metric d on $S \times T$ one can use $d((s, t), (s', t')) = d_S(s, s') + |t' - t|$ (or the maximum) where d_S is any metric on S such that the neighbourhoods $N(s)$ have bounded diameter.

This is a stronger definition than used in [10] but still holds for many appropriate CML with symbolic dynamics, including our Examples 1–4. For these examples, let $\lambda = \frac{2}{2-\varepsilon}$ and choose $D_0 > 2\frac{\log 1/\delta}{\log \lambda}$. Let B_i be the $D_0/2$-neighbourhood of A_i. Let σ be the symbolic state of $x(i)$ on each B_i, extended arbitrarily on the rest of $S \times T$. Let y be the orbit corresponding to σ. Each y_s^t is determined by $\Sigma_s^t = (\sigma_r^u)_{r \in N(s), u \geq t}$, because for any $\tau \geq t$,

$$y_s^t = g_t(\ldots(g_\tau(x_s^{\tau+1})\ldots)$$

where g_u denotes the inverse branch of the local map corresponding to the symbolic state on $N(s)$ at time u. The g_u are contractions by a factor of at least λ, so by taking $\tau \to \infty$ we obtain a unique y_s^t for each choice of Σ_s^t, and all those with given σ on $N(s) \times [t, \tau]$ are contained in an interval of length at most $\lambda^{t-\tau}$. In particular, y is determined to within δ on each site of A_i by σ on B_i, so agrees with $x(i)$ to within δ on each site of A_i.

The above definition of indecomposable can also be applied to PCA by replacing "orbit" by "allowed realisation" and dispensing with δ (it is similar to the "strong irreducibility" of [6]), though probabilists often ask for a stronger property to rule out events with vanishing probability (see the literature on "positive rates", e.g. [12]).

2.4 CML with Eternal Transients

An interesting phenomenon discovered numerically by [8] and again by Livi et al. (see [11] for recent work of theirs and references), is that some spatially extended systems with a small attracting subset can take extraordinarily long times to discover the attractor; instead they bounce around for a long time

in a chaotic transient. This suggests that for the infinite system there is an additional pure phase to that on the attracting subset. I present here a proved example of this.

Example 5. Let $S = \mathbb{Z}$ and $N(s) = \{s, s+1\}$ for each $s \in S$. For $x_s \in [0,1]$ define $\sigma_s = \pm$ according as $x_s \geq \frac{1}{2}$ or $< \frac{1}{2}$. Then apply the local maps indicated in Fig. 5, where the slopes are $\frac{1}{1-\varepsilon}$ and $\frac{1}{\varepsilon}$, for ε small enough.

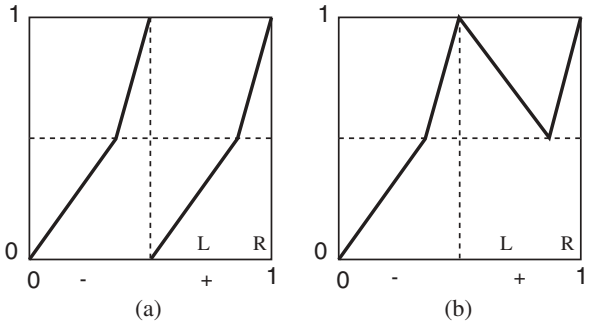


Fig. 5. Maps defining a CML with eternal transient phase for ε small enough (drawn with $\varepsilon = \frac{1}{4}$), to be applied when the right-hand neighbour is (**a**) $-$, (**b**) $+$

The subset $[\frac{1}{2}, 1]^S$ is invariant, so there is a Gibbs phase (in fact pure and corresponding to the Bernoulli measure $B(1-\varepsilon, \varepsilon)$ on it), and it would seem that it ought to absorb the orbits of most initial conditions. This is indeed the case if ε is not too small, but remarkably for ε small enough there is another pure Gibbs phase, supported on the complement.

The construction is based on *Stavskaya's PCA* (a case of the discrete-time asymmetric contact process, also known as directed site percolation) on $\{+,-\}^S$ where

$$P(\sigma_s^{t+1} = + \mid \sigma^t) = \begin{cases} 1 & \text{if } \sigma_s^t = \sigma_{s+1}^t = + \\ \varepsilon & \text{otherwise} \end{cases}$$

(the remaining probability $1-\varepsilon$ in the second case going to $\sigma_s^{t+1} = -$) for which corresponding behaviour has been proved if ε is small enough (≤ 0.09, though numerically it is good up to about 0.31 [29]). To make the construction fit within our general framework of expanding maps with symbolic dynamics, we subdivide the $+$ interval at $1-\varepsilon/2$ into two subintervals $L \cup R$ and use a map that expands each of L and R to $[\frac{1}{2}, 1]$ for the $++$ case, so that we obtain symbolic dynamics with alphabet $\{-, L, R\}$. The same phenomenon would occur, however, if the map of $[\frac{1}{2}, 1]$ in the $+$ case were replaced by any other with uniform density as mixing measure.

It would be interesting to analyse examples where the map of $[\frac{1}{2}, 1]$ in the $+$ case has a globally attracting fixed point, to exhibit the phenomenon of "stable chaos" of [11].

Many variants of this example can be constructed, using more general discrete-time contact processes or PCAs associated to directed percolation, though I am not aware of proofs of non-unique phase for most of these (there is a proof for a broad class of continuous-time contact processes, e.g. [21]).

2.5 Pre-indecomposability

The eternal transient example is not indecomposable, since it possesses the forward invariant subset $[\frac{1}{2}, 1]^S$. In [10] we got round this by allowing D_0 to depend on the diameters of the A_i, which allows us to wait for influence of distant $-$ to propagate in and exploit the discontinuity, but I think that obscured the point that really the CML should be considered only "pre-indecomposable" and it is not appropriate to ask for more.

I'll explain the concept first in the context of Markov chains. I say a Markov chain is *pre-indecomposable* if it has an aperiodic irreducible component which can be reached from anywhere. For a finite-state Markov chain this is a necessary and sufficient condition for unique phase.

Similarly, I'll say a dynamical system is *pre-indecomposable* if it has an attractor A with the specification property and for all $\delta > 0$ there exists $\tau > 0$ such that for every orbit up to time 0 and orbit on A from time τ there exists one within δ of each. Thus we can include the basin of the attractor and furthermore the closure of the basin. For $C^{1+\alpha}$ systems in finite dimensions I think this suffices for unique phase (Bowen's fat horseshoe example shows that the Hölder condition can not be dropped).

Thus I'll say a CML is *pre-indecomposable* if it has an indecomposable invariant subset A and for all $\delta > 0$ there exists τ such that for every orbit x up to time 0 and orbit y on A from time τ onwards there exists an orbit z within δ of each.

This holds for Example 5: take $A = [\frac{1}{2}, 1]^S$ and $\tau > 2\frac{\log 1/\delta}{\log \lambda}$ where $\lambda = \frac{1}{1-\varepsilon}$, then let σ equal that for x up to time $\tau/2$ and that for y thereafter, and let z be the orbit with symbol table σ, which exists because the only forbidden transition is $++ \to -$.

2.6 Construction of Invertible Examples

So far, the examples have been non-invertible. This would not disturb researchers in dissipative PDE (like reaction-diffusion systems), but for other contexts like nonlinear wave equations or oscillator networks, it would be better to have invertible examples.

A simple way to make an invertible CML to simulate a PCA is to replace the expanding 1D maps by "baker's maps": take local state space M to be

$Q = [0,1]^2$ with coordinates (x, y), and construct a map which on the x coordinate is as before and on the y coordinate is just (for 2-state PCA)

$$y' = \begin{cases} y/2 & \text{for } x < 1/2, \\ (1+y)/2 & \text{for } x \geq 1/2\ . \end{cases}$$

The Gibbs phases for this CML are just those for the expanding map examples in x, supplemented by the distribution in y corresponding to the possible symbol histories of x, converted to binary expansions. For example, for G^+ in Example 1, for any (s,t) the marginal on x_s^t has a step function density with values $c(\varepsilon), 1-c(\varepsilon)$ for $x > \frac{1}{2}$, $x < \frac{1}{2}$, and the marginal on y_s^t is a singular continuous distribution because the symbol histories form a source with entropy (in base-2) $h < 1$, thus for large N most of the probability is supported on about 2^{Nh} of the 2^N intervals $[\frac{p-1}{2^N}, \frac{p}{2^N}]$ for $p = 1, \ldots, 2^N$, whose total length $2^{N(h-1)} \to 0$ as $N \to \infty$, yet there are no atoms.

Note that in general for an invertible system one obtains different Gibbs potentials for forward and backward evolution (see Sect. 3.8), and hence in general different sets of Gibbs phases. For the invertible CML just constructed, there is a unique backwards Gibbs phase and it is uniform in y.

An alternative construction, for which the local maps are continuous (though the CML still has discontinuities with respect to coupling), is to take piecewise affine continuous deformations of hyperbolic automorphisms of a 2-torus, with continuously deformed Markov partition, but there are restrictions on the possible transition probability matrices, so I won't pursue it.

2.7 Reversible CML with Non-reversible Phases

In some contexts (e.g. much of Hamiltonian mechanics), the dynamics is not only invertible but also reversible. A dynamical system is *reversible* if it is invertible and conjugate to its inverse via an involution (a map R on the state space whose square is the identity). So for any motion there is an equally possible reversed motion. The prime example is celestial mechanics: if one were to reverse the momenta of all the planets one would obtain the time-reversed motion.

The basic laws we use for the universe are reversible, yet we see a very definite arrow of time. This is expressed in various ways like the second law of thermodynamics, cause precedes effect, and outgoing radiation conditions. Why are only motions compatible with these conditions apparent and not their time-reverses?

The prevailing view (e.g. [25]) attributes the arrow of time to a very special choice of initial condition (big bang), but I consider that unsatisfactory as an explanation, because it is still locked in the "cause precedes effect" mindset. In contrast, I suggest it is a result of non-unique space-time Gibbs phase at some level, probably quantum gravity, selecting one light-cone in preference to the other in large space-time patches.

Here I make a toy model to illustrate the possibility. I use chaotic dynamics instead of quantum physics to provide the stochastic effect (as done in [2] for other purposes).

The starting point is synchronous 2D kinetic Ising models. These are PCA (of period-2 rather than autonomous, but that does not cause essential trouble) which are reversible (see the Appendix for a discussion of what it means for a PCA to be reversible) and which have precisely two pure phases. I'll simulate this by a reversible CML. The phases are reversible but by making a suitable skew product over this CML, I'll then make a reversible CML with non-reversible phases.

For kinetic Ising model, take state space $\{+,-\}^S$ with $S=\mathbb{Z}^2$ and define the neighbourhood $N(s)$ of $s\in S$ to consist of $s\pm(1,0)$ and $s\pm(0,1)$. Choose $J\in\mathbb{R}_+$. Say a space-time site (s,t) is "even" or "odd" according as s_1+s_2+t is even or odd. For (s,t) even, let $m=\sum_{r\in N(s)}\sigma_r^t$ and "Glauber dynamics"

$$P(\sigma_s^{t+1}=\pm|\ \sigma^t)=\frac{1}{2}\left(1\pm\tanh(Jm)\right)=\frac{\exp\pm Jm}{\exp Jm+\exp -Jm}$$

and for (s,t) odd, let $\sigma_s^{t+1}=\sigma_s^t$. This PCA has the nice features that it is reversible and its Gibbs phases are in one-to-one correspondence with those for the equilibrium statistical mechanics of the 2D Ising model at temperature 1, with potential $-\sum_s\frac{J}{2}\sigma_s\sum_{r\in N(s)}\sigma_r$ (the factor $\frac{1}{2}$ is to compensate for counting each bond twice). For J large enough, the 2D equilibrium Ising model has two pure phases, thus so does the kinetic Ising model. There are other transition rates that achieve the same properties (e.g. Metropolis rates). Also the kinetic Ising model can be reformulated as an autonomous PCA by going to a diagonally moving frame (Ex 1.4b of [29]), so one could produce an autonomous CML to simulate it, but the moving frame complicates the reversibility properties.

The kinetic Ising model can be simulated by an expanding CML using the construction of Sect. 2.2 again. Indeed, a variant of the method was used by Sakaguchi [27] long before our work, but with a variable partition and constant slope on each partition element instead of fixed partition and varying the slopes of the map within each partition element, which to my mind makes the non-uniqueness of phase less marked because both phases have uniform marginal density on all finite subsets. Also, it is not clear whether the paper contains a complete proof, and the construction is less flexible than ours. Finally, Sakaguchi's CML is second-order, meaning that to determine x^{t+1} one needs both x^t and x^{t-1} (though through only the symbolic state of x^{t-1}), so to make a first-order CML out of it, one would have to augment the local state, which becomes messy.

We want to simulate the kinetic Ising PCA by a reversible CML, however. The following variant of our baker's map construction produces one.

Example 6. Take $S=\mathbb{Z}^2$, local state space $M=[-1,0]^2\cup[0,1]^2$ (really, I mean disjoint union, so there are two distinct points $(0,0)$) with coordinates

(x, y), for each $s \in S$ let $N(s) = \{s \pm (0,1), s \pm (1,0)\}$, and let $\sigma_s = \pm$ according as $(x_s, y_s) \in [0,1]^2$ or $[-1,0]^2$. Choose $J \in \mathbb{R}_+$. For (s,t) even, evaluate $m(s) = \sum_{r \in N(s)} \sigma_r \in \{-4,-2,0,2,4\}$ and $p_\pm = \frac{1}{2}(1 \pm \tanh Jm)$, and apply the following map to (x_s, y_s) (see Fig. 6):

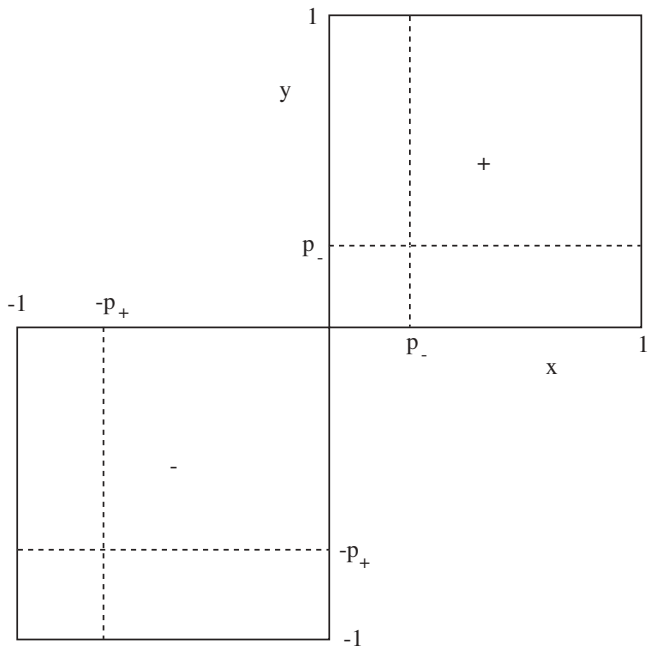


Fig. 6. The domains (*vertical rectangles*) and ranges (*horizontal rectangles*) of the affine pieces of the local maps for Example 6

$$\text{For } 0 < x < p_- \quad \begin{cases} x' = -\frac{x}{p_-} \\ y' = -p_+ y \end{cases}$$

$$\text{For } p_- < x < 1 \quad \begin{cases} x' = 1 - \frac{1-x}{p_+} \\ y' = 1 - p_+(1-y) \end{cases}$$

$$\text{For } -p_+ < x < 0 \quad \begin{cases} x' = -\frac{x}{p_+} \\ y' = -p_- y \end{cases}$$

$$\text{For } -1 < x < -p_+ \quad \begin{cases} x' = -1 + \frac{1+x}{p_-} \\ y' = -1 + p_-(1+y) \end{cases}.$$

For (s,t) odd, apply the identity map.

This CML is reversible with respect to $R : x \leftrightarrow y$ and $t \mapsto 1 - t$ (as the CML is non-autonomous, the notion of reversibility needs a slight modification, hence the necessity to specify which reflection to take in time). It is also

indecomposable (similar proof to Example 1). For $J > J_c = \log(1+\sqrt{2})$, there are $c > \frac{1}{2}$ (in fact, $c = \frac{1}{2}(1+(1-1/\sinh^4 J)^{\frac{1}{8}})$) and precisely two pure phases $G^{\pm}$, and $G^{\pm}(\sigma_s^t = \pm) = c$. Both phases are reversible, however, because they are essentially just the Gibbs phases of the kinetic Ising PCA, which are reversible.

To make a reversible CML with non-reversible Gibbs phases, I first note that Example 6 is also reversible with respect to $R' : (x, y) \mapsto (-y, -x)$, $t \mapsto 1 - t$, and $G^{\pm}$ are not reversible with respect to this involution. I am reluctant to claim this as a valid example of reversible CML with non-reversible phases, however, because one could say I just chose the wrong involution. So I seek to modify Example 6 to lose R-reversibility but keep R'-reversibility. This can be done by a skew-product over Example 6, where dynamics in additional variables z is added.

Example 7. In addition to the (x, y) dynamics of Example 6, choose an invertible map g of a manifold Z which is not reversible with respect to any involution (this rules out Z being finite), take local state space $([0,1]^2 \cup [-1,0]^2) \times Z$ with coordinates (x, y, z), at odd (s, t) take $z' = g(z)$ if $\sigma = +$ and $z' = g^{-1}(z)$ if $\sigma = -$, and at even (s, t) take $z' = z$. This is reversible with respect to $R'' : (x, y, z) \mapsto (-y, -x, z)$ but not with respect to any involution which interchanges x and y. If g has unique and distinct forward and backward phases, then the resulting CML is indecomposable (because given $\delta > 0$ one can insert a bounded band of symbols $+$ or $-$ to move z_s from one phase to the other), and has pure phases corresponding to $G^{\pm}$ on (x, y) and the forwards (respectively backwards) phase of g, because with G^+-probability 1, for each site s, $\sum_{t=0}^{T} \sigma_s^t \to +\infty$ as $T \to +\infty$, so z_s^t moves to the forward phase of g as t increases, and similarly for G^-. These phases are not reversible.

Concretely, one can take $Z = [-1, +1]$ and g to be any diffeomorphism of $[-1, +1]$ with fixed point set $\{-1, +1\}$ and the product of the slopes at the ends not equal to 1. The final condition rules out existence of any smooth reversor, but one might prefer to rule out all continuous reversors. Then one could take something like $Z = \mathcal{B}^3$, the unit ball in three dimensions, and g to be a Bernoulli diffeomorphism on its boundary S^2 (which can be made by smoothing a pseudo-Anosov map of S_4^2, the sphere with four punctures, following Katok) and to attract all the interior to a fixed point at the origin. Alternatively, one could take $Z = \mathbb{T}^2$ and g a non-reversible uniformly hyperbolic map: non-reversibility can be achieved by making the forward and backward SRB measures not mappable onto each other.

Although Example 7 solves the problem posed, it is not ideal. Firstly, it leaves the realm where Gibbs phases are defined, since the skew products are not hyperbolic in any good sense: the z dynamics is a sequence of g and g^{-1}, so any amount of expansion in Z can be undone by a suitable symbol sequence

in the future. Secondly, my original aim was to make examples with phases carrying a thermodynamic arrow of time or selecting a radiation condition, but z has no interpretation as entropy and there is no coupling of the z dynamics between sites. Perhaps a thermodynamic arrow of time could be achieved by a space-time analogue of the interface phases of the 3D Ising model, and a radiation condition by skew-product with a suitable deterministic cellular automaton.

An interesting question is whether examples of reversible CML with non-reversible phases can be made that are also symplectic, because that would suggest that non-reversible phase could occur in spatially extended Hamiltonian systems.

3 Motivation for Gibbs Phases

3.1 Probabilistic Behaviour

For a deterministic dynamical system, each initial condition determines a unique trajectory, but it often happens that many initial conditions lead to similar trajectories as time goes to infinity. This holds in a strong sense for initial conditions in the basin of an attracting fixed point. It holds in a slightly weaker sense for initial conditions in the basins of different points of the same attracting periodic orbit, because their orbits become similar only if one allows to compare suitable time-shifts. For chaotic systems (meaning at this stage that they have sensitive dependence on initial conditions), trajectories of typical pairs of initial conditions are unlikely to converge together, but the trajectories of many initial conditions may still be similar to each other in a statistical sense. The tidiest way to say this is that the probability distribution of the state at time t with respect to any member of a specified class of measures on initial conditions may converge in weak-* as $t \to \infty$ to a common probability measure (a sequence μ_n of measures on a topological space M *converges in weak-** if for all continuous functions $\phi : M \to \mathbb{R}$ the sequence of expectations $\mu_n(\phi)$ converge; the terminology comes from viewing measures as a subset of the dual of the space of continuous functions on M).

For example, for $f : x \mapsto 4x(1-x)$ on $[0,1]$ and any absolutely continuous measure μ on $[0,1]$, i.e. one which gives probability zero to all sets of Lebesgue measure zero, then μ is represented by a density $\varrho \in L^1[0,1]$ with $\int \varrho(x)dx = 1$. Define the "transfer" operator T on densities ϱ by $\int g(x)(T\varrho)(x)dx = \int g(f(x))\varrho(x)dx$ for all $g \in L^\infty[0,1]$; then $T^n\varrho \to \bar{\varrho}$ as $n \to \infty$ where

$$\bar{\varrho}(x) = \frac{1}{\pi\sqrt{x(1-x)}}$$

meaning $\int g(x)(T^n\varrho)(x)dx \to \int g(x)\bar{\varrho}(x)dx$ for all $g \in L^\infty[0,1]$. Note that it follows that for any $\tau > 0$, the probability distribution for $(x^t, x^{t+1}, \ldots, x^{t+\tau})$ converges as $t \to \infty$ (in weak-* topology) to that generated by f from $\bar{\varrho}$.

I think about this is in terms of the orbits of uncertain initial conditions. If one does not have infinite precision over the initial condition then it is more reasonable to consider the initial condition as a probability measure, perhaps tightly peaked around some point but nonetheless absolutely continuous. This point of view allows me to say that the orbits of all uncertain initial conditions have the same statistics in the above example, whereas all I can claim for exact initial conditions is that almost all lead to the same statistics: it is essential to remove a set of Lebesgue measure zero of exceptional orbits, like those which converge to an unstable fixed point or periodic orbit.

One can make examples, however, with a two-band attractor or more than one attractor, for which such convergence does not happen unless one starts in the same basin. Such examples have more than one probabilistic behaviour.

I call probability measures on realisations in time originating from nice enough initial measures far in the past *phases* (not to be confused with phase in the sense of position in a cycle or in a bundle over a circle). They are sometimes called "states" but I reserve this word for the points of the state space of the dynamics (confusingly called "phase space" in the older literature!). Their time-averages are often called natural or physical or SRB measures.

Any convex combination of phases is also a phase: for example start with initial probability distribution supported partly in the basin of one attractor and partly in that of another, but the interest focusses on the *pure phases*, which are the phases which can not be expressed as a convex combination of others (sometimes called *extremal phases*).

3.2 Ordinary Symbolic Dynamics

Some dynamical systems have the nice feature that there is a partition of the state space into regions and a directed graph between the regions such that every path in the graph (called a *symbol sequence*) occurs for some orbit and there is precisely one orbit for each path in the graph. Thus to each symbol sequence σ there corresponds an initial condition $x^0(\sigma)$. The regions are allowed to overlap along "negligible" sets, so orbits which pass through such overlap sets may have more than one symbol sequence, but this is a relatively unimportant issue. The system is said to have *symbolic dynamics*.

Symbolic dynamics occurs for every locally maximal compact uniformly hyperbolic set for a C^1 diffeomorphism, and also for suitably constructed maps with jump discontinuities, like expanding maps for which each interval of continuity is mapped onto a union of intervals of continuity.

3.3 Gibbs Phases for Markov Chains

Before considering the asymptotic probabilistic behaviour of deterministic dynamical systems with symbolic dynamics, let's study the same question in the simpler context of a Markov chain.

It is usual to specify a Markov chain by giving the forward transition probabilities $p_{\sigma\sigma'}$ from states σ to σ', which define probabilities

$$P\{\sigma^1, \ldots, \sigma^\tau |\ \sigma^0\} = \prod_{t=0}^{\tau-1} p_{\sigma^t \sigma^{t+1}}$$

for the initial value problem. Actually, to avoid confusion with the official definition of Markov chain, which specifies probabilities for all sequences not just conditional on initial conditions, I should give this a different name, like "pre-Markov chain". The distinction is important because we are trying to decide when specification of the transition probabilities implies a unique probability distribution on doubly infinite sequences. Nevertheless, I'll follow many physicists and call it a "Markov chain" except in the Appendix.

Even though they do not determine the unconditional probabilities of sequences, the transition probabilities do define the probability distributions for the two-point boundary value problems with σ^m, σ^n given (assuming there exists a path with positive probability):

$$P\{\sigma^{m+1}, \ldots, \sigma^{n-1} |\ \sigma^m, \sigma^n\} = \frac{1}{Z} \prod_{t=m}^{n-1} p_{\sigma^t \sigma^{t+1}}$$

where $Z = Z(\sigma^m, \sigma^n)$ is a normalisation constant. This can be rewritten as

$$P\{\sigma^{m+1}, \ldots, \sigma^{n-1} |\ \sigma^m, \sigma^n\} = \frac{1}{Z} \exp\left(- \sum_{t=m}^{n-1} \log \frac{1}{p_{\sigma^t \sigma^{t+1}}} \right) .$$

Thus $\log \frac{1}{p_{\sigma\sigma'}}$ can be interpreted as an "energy" contribution from the transition $\sigma \to \sigma'$ of state along a 1D statistical mechanics chain at temperature 1, and the above equation for the probability distributions for two-point boundary value problems is precisely that defining (Boltzmann-)Gibbs phases of such a statistical mechanics model.

A *Gibbs phase* for a Markov chain is a probability distribution on doubly infinite sequences of states such that the conditional probability distribution for all two-point boundary value problems is given by this formula. Of course, for an aperiodic irreducible Markov chain there is a unique Gibbs phase, generated by its stationary measure and the transition probabilities.

3.4 Gibbs Phases for Expanding Maps

For expanding maps with symbolic dynamics and Hölder continuous derivative on each branch there is an analogous way to determine the possible long-time statistics of orbits, but in general we have to consider boundary value problems where all the symbols before some time and after some other time are given, rather than just two-point.

A map f is *expanding* if there is a $\tau > 0$ and $\lambda > 1$ such that f^τ expands the distance between any two nearby points by at least λ. Note that such maps are necessarily noninvertible, but I'll consider invertible maps later.

If we start from an absolutely continuous measure with density ϱ at some large negative time M, and we observe the symbol sequence σ from time M to some large positive N, then we could ask what is the probability of the subsequence $\sigma^m, \ldots, \sigma^n$ for some fixed times m and n between M and N, given the remaining symbols. I will find it useful to assume that $\log \varrho$ is Hölder continuous on each subset with a given long enough symbol sequence, though this should not really be necessary.

The answer is proportional to $\int \varrho(x^M)\, dx^M$ over the set of x^M with symbol sequence $\sigma^M, \ldots, \sigma^N$, this being the measure of the set of initial conditions which follows the given symbol sequence. Using the map f, which is invertible in the domain where the given symbol sequence occurs, we can change variable of integration from x^M to $x^{M+1} = f(x^M)$, obtaining

$$\int \varrho(x^M)\, dx^M = \int \frac{\varrho(x^M)}{|\det Df_{x^M}|}\, dx^{M+1}$$

over the set with symbol sequence $\sigma^{M+1}, \ldots, \sigma^N$, where x^M is regarded as a function of x^{M+1} by using the appropriate inverse branch of f. Iterating this procedure up to any time $K \leq N$, we find that the probabilities of symbol sequences $\sigma^m, \ldots, \sigma^n$ given the rest are proportional to

$$\int \frac{\varrho(x^M)}{\prod_{t\in\{M,\ldots,K-1\}} |\det Df_{x^t}|}\, dx^K$$

over the set with symbol sequence $\sigma^K, \ldots, \sigma^N$.

For $K > n$, the range of integration is independent of $\sigma^m, \ldots, \sigma^n$. For M large (negative), the set of x^M with given symbol sequence $\sigma^M, \ldots, \sigma^{m-1}$ has exponentially small diameter in $m - M$, so by the assumed continuity of $\log \varrho$, ϱ can be approximated by constant density on it. If we choose K to be roughly halfway between n and N, then for N large, $\prod_{t\in\{M,\ldots,K-1\}} |\det Df_{x^t}|$ is close to constant for each symbol sequence (because Df is Hölder continuous and the sets over which the x^t can range are exponentially small in $N - t$). Thus as $M \to -\infty$, $N \to +\infty$, the conditional probabilities of $\sigma^m, \ldots, \sigma^n$ are proportional to $1/\prod_{t\in\{M,\ldots,K-1\}} |\det Df_{x^t}|$, which acquires a well defined meaning because the effect of the terms far from times $t \in \{m, \ldots, n\}$ depends exponentially weakly in $K - t, t - M$ on $\sigma^m, \ldots, \sigma^n$.

It is common to rewrite this relative conditional probability as

$$\exp\left(-\sum_{t\in\{M,\ldots,K-1\}} \log|\det Df_{x^t}|\right)$$

to make it look like the condition for a Gibbs measure in statistical mechanics. Thus we see that the possible probabilistic behaviours for expanding maps

with symbolic dynamics started far in the past from an absolutely continuous measure correspond to the Gibbs phases for a 1D spin chain with energy contribution from site t given by $\log|\det Df_{x^t}|$, where x^t is the state at time t corresponding to a given symbol sequence. Since x^t depends exponentially weakly on symbols distant from t and Df is Hölder continuous, this gives an exponentially decaying interaction between sites in the spin chain.

3.5 Gibbs Phases for PCA

For a PCA with space S, if we wish to know the probability measures on all possible realisations on $S \times T$ (with $T = \mathbb{Z}$) which are consistent with the transition probabilities, including arbitrarily far in the past, then we can again write them as Gibbs phases.

As for general Markov chains, we consider boundary (rather than initial) value problems, but now in space-time rather than just time. For each bounded subset Λ of space-time $S \times T$ and state η on the complement of Λ, we can ask for the conditional probability distribution for the state σ on Λ. Define the *templates* for the PCA to be the subsets of $S \times T$ of the form $\{(s, t+1)\} \cup \{(r, t) : r \in N(s)\}$, where $N(s)$ is the neighbourhood of dependence for s, and for state σ and template τ define the *weight* $w_\tau(\sigma)$ to be the transition probability $p_\tau(\sigma)$ for the transition represented by the restriction of σ to τ. Then

$$P(\sigma \text{ on } \Lambda|\ \eta \text{ off } \Lambda) = \frac{1}{Z} \prod_{\tau \cap \Lambda} p_\tau(\sigma \vee \eta)$$

where $\tau \cap \Lambda$ means the set of templates having nonempty intersection with Λ, $\sigma \vee \eta$ is the state on $S \times T$ given by σ on Λ and η off Λ, and Z is chosen to normalise the distribution (provided that there is at least one σ compatible with η). Thus we view PCA on S as a class of Markov random field on $S \times T$ (with the special features that the templates have the shape above and the weights on a template with given state on $N(s)$ sum to 1 over the state at the "apex"). This conditional probability can be rewritten as

$$P(\sigma \text{ on } \Lambda|\ \eta \text{ off } \Lambda) = \frac{1}{Z} \exp\left(- \sum_{\tau \cap \Lambda} \log \frac{1}{p_\tau(\sigma \vee \eta)} \right)$$

which looks like the condition for a Gibbs distribution for a statistical mechanics spin lattice on $S \times T$ where each template τ contributes an energy $\log \frac{1}{p_\tau(\sigma)}$ (in temperature units).

3.6 CML with Symbolic Dynamics

The concept of symbolic dynamics extends to some CML, in particular those we constructed to simulate PCA, and also uniformly hyperbolic CML as I'll

indicate below. For a CML, a trajectory x is an array of x_s^t with $s \in S, t \in T$ ($T = \mathbb{Z}_+$ or $\mathbb{Z}$, for the expanding or invertible case, respectively), and the symbol sequence is replaced by a symbol table $\sigma = (\sigma_s^t)_{s\in S, t\in T}$.

Our examples have constant derivative on each piece corresponding to a given symbolic transition, but for general expanding CML with symbolic dynamics, the derivative depends on the exact state making a given symbolic transition, which is determined by more distant symbols in space and time. So to justify the concept of Gibbs states for general expanding CML with symbolic dynamics, we need to examine the dependence of the derivative on the full symbol table. Following [26, 23], here is a sketch (which applies equally well in the invertible case).

For CML F on M^S, the orbits $x^{t+1} = F(x^t)$, $t \in T$, correspond to the fixed points of the map $\mathcal{F}$ on $M^{S\times T}$ defined by

$$(\mathcal{F}x)_s^t = F_s(x^{t-1}) \ .$$

If F is differentiable (in uniform norm) with bounded derivative, then so is $\mathcal{F}$. I say an orbit x is *uniformly hyperbolic* if it is a non-degenerate fixed point of $\mathcal{F}$, i.e. $I - D\mathcal{F}_x$ is invertible; since $D\mathcal{F}$ is bounded it follows that $(I - D\mathcal{F}_x)^{-1}$ is bounded. I say a set U of uniformly hyperbolic orbits is *uniformly hyperbolic* if there exists $C \in \mathbb{R}$ such that $\|(I - D\mathcal{F}_x)^{-1}\| \le C$ for all $x \in U$.

If the coupling in F is local (or exponentially decaying in a suitable sense) then it can be shown that this definition of uniform hyperbolicity is equivalent to the usual one in terms of a splitting of the tangent space into a contracting subspace and a backwards contracting subspace with uniform exponential contraction estimates and angle of splitting bounded away from zero. In addition, however, we obtain exponential decorrelation in space. The point is that uniform hyperbolicity plus exponentially decaying coupling implies that the matrix elements of $(I - D\mathcal{F}_x)^{-1}$ decay exponentially from the diagonal in both space and time. In particular, this implies that the dependence of x_s^t on symbol $\sigma_{s'}^{t'}$ decays exponentially with $d((s,t),(s',t'))$, which will be required in the next section.

This viewpoint on uniform hyperbolicity for CML has many consequences for the dynamics at the topological level as well as probabilistic. In particular, it allows one to construct symbolic dynamics for weakly coupled uniformly hyperbolic maps [26] and to obtain many of the results of Afraimovich's lectures. The idea is that given a uniformly hyperbolic set U of orbits for CML F then all C^1-small perturbations of F possess a unique continuation of U, with topologically equivalent dynamics. Specifically, if F depends on parameters λ, the continuation $x(\lambda)$ of an orbit $x(0) \in U$ is given by

$$\frac{\partial x}{\partial \lambda} = (I - D\mathcal{F}_x)^{-1} \frac{\partial \mathcal{F}}{\partial \lambda}(x) \ .$$

This can be used to make further examples of interesting topological dynamics in CML. For example, one can make CML with localised chaotic attractors, by

taking the on-site dynamics to have an attracting fixed point and a uniformly hyperbolic chaotic attractor (e.g. Plykin or solenoid), choosing one site to be on the chaotic attractor and the rest on the fixed point, and adding any smooth weak coupling [23].

Furthermore, the shadowing theorem extends to this space-time context, so the approach allows to construct symbolic dynamics for any locally maximal uniformly hyperbolic set for a CML (choose a sufficiently fine set of time-periodic orbits and then use shadowing to associate to any orbit at each $(s,t) \in S \times T$ one of the periodic orbits, thus constructing its symbol table).

3.7 Gibbs Phases for Expanding CML with Symbolic Dynamics

For a CML F on M^S, the phases should describe the statistics of orbits from all nice enough initial probability distributions on the state space, started far enough in the past. It is not yet clear what is the sensible class of initial probability distributions to consider: see Jarvenpää's lectures. For safety, I will restrict attention to those with a C and a $\delta > 0$ such that their marginals on any finite subset $L \subset S$ have an L^1 density ϱ_L and

$$|\log \varrho_L(x) - \log \varrho_L(y)| \leq C \sum_{s \in L} d(x_s, y_s)^\delta$$

as in [4] and [10], but in fact I will require it only when x and y have the same symbol table on some neighbourhood of $L \times \{0\}$ in $S \times T$. I believe this Hölder condition on the logarithm of the densities is unnecessarily strong, but I have not yet formulated a better one.

Given an initial measure satisfying the Hölder condition at a large negative time M, we ask for the probabilities of symbol tables on a given $\Lambda \subset S \times T$, conditional on the symbol table outside Λ. Fatten Λ to some larger patch Λ', in both space and time, to allow for the fact that in general the symbols on Λ depend on the state off Λ as well as on Λ, but exponentially weakly as the distance away from Λ increases. Because the CML has finite range coupling, disturbances can propagate at most at some speed. Let S^M be the subset of S at time M capable of influencing the state on Λ'. Fatten S^M in space to a set J. Then the conditional probabilities P of symbols on Λ are proportional to $\int \varrho_J(x^M) \prod_{s \in J} dx_s^M$ over the subset of states x^M on J giving rise to the given symbol table (modulo a slight dependence on the state outside J in general, which one should bound).

Change variable from x_s^M to x_s^{M+1} at all sites $s \in S^M$ such that $N(s) \subset S^M$, using the map F. Call this set of sites S^{M+1}. The volume change factor is the determinant of the block of DF corresponding to S^{M+1}, which I'll denote $DF_{S^{M+1}}$, evaluated at the state x^M at time M. Then

$$P = \frac{1}{Z} \int \frac{\varrho_{S^M}(x^M)}{|\det DF_{S^{M+1}}(x^M)|} \prod dx_s^t$$

where (s,t) ranges over S^{M+1} at time $M+1$ and the remaining sites of J at time M.

It is convenient to rewrite

$$|\det DF_{S^{M+1}}(x^M)|^{-1} = \exp\left(-tr\log DF_{S^{M+1}}(x^M)\right).$$

For any invertible operator A, $tr\log A$ can be calculated in an arbitrary bounded connected neighbourhood of a given A_0, up to a constant, by writing $A = A_0\prod_i(I-B_i)$ for a suitable finite product of operators with each $||B_i|| < 1$, using $\log(I-B) = -\sum_{n\geq 1}\frac{B^n}{n}$ so $tr\log(I-B) = -\sum_{n\geq 1}\frac{trB^n}{n}$, and letting $tr\log A = \sum_i tr\log(I-B_i)$ (it is not necessary to add a value for $tr\log A_0$ because the choice would change only the normalisation constant Z). Alternatively, it can be calculated as $\int_0^1 tr\left(A^{-1}(\lambda)A'(\lambda)\right)\,d\lambda$ along any differentiable path of invertible operators to A from a convenient reference point A_0. Note that neither approach defines $\log A$ itself if the B_i do not commute or $A'(\lambda)$ does not commute with $A(\lambda)$ along the path (one has to use the Campbell-Baker-Hausdorff formula; we neglected to mention this in [10]), but both always give the correct trace.

Next, we can expand

$$\exp\left(-tr\log DF_{S^{M+1}}(x^M)\right) = \exp\left(-\sum_{s\in S^{M+1}} tr(\log DF)_{ss}(x^M)\right)$$

where $(\log DF)_{ss}$ denotes the diagonal block of $\log DF$ corresponding to site s (of dimension equal to that of the local state space, so 1 for all our expanding examples). Since all that matters is the sum over s, one can again replace the true $\log DF$ by either of the surrogates of the previous paragraph.

Now iterate the change of variables to a nested sequence of subsets S^t at times $t = M,\ldots,\tau$, where τ is the largest time in Λ'. Then

$$P = \frac{1}{Z}\int \varrho_{S^M}(x^M)\exp\left(-\sum_{(s,t)\in K} tr(\log DF(x^t))_{ss}\right)\prod_{(s,t)\in L} dx_s^t$$

where K is the subset of $S\times T$ with $t\geq M$ that can influence the state on S^τ at time τ and L is the union of its "boundary" and the sites of J at time M not in S^{M+1}. See Fig. 7 for an illustration for the case of a 1D nearest neighbour CML.

Since the dynamics is expanding and Λ is contained in the subset of $S\times T$ which can influence the state on L, the domain of integration is the same for all σ (given η). Now x on J determines x on a fattening K' of K. The exponent is almost independent of x in the domain of integration for each fixed σ and η, because the symbolic state determines the actual state with exponential accuracy and DF is assumed to be Hölder continuous. Finally, ϱ_{S^M} can be approximated by a constant, independent of σ, because of the Hölder condition on initial measures.

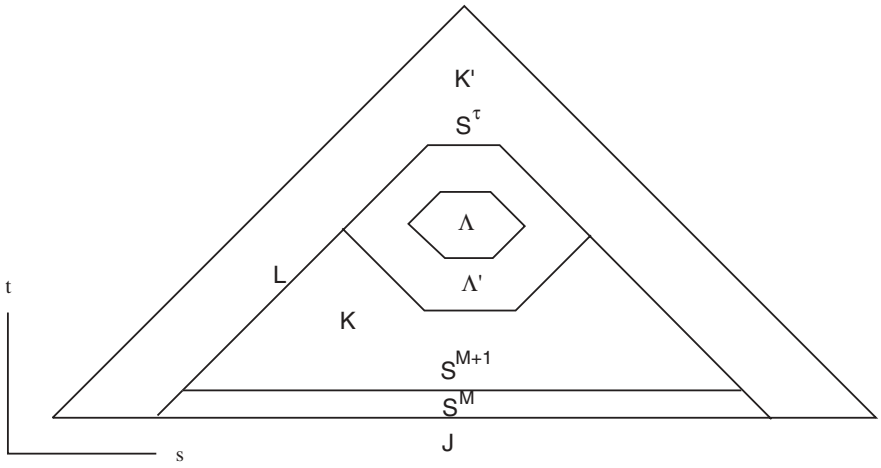


Fig. 7. Illustration of the change of variables for a 1D nearest neighbour CML

Thus, taking appropriate limits,

$$P(\sigma \text{ on } \Lambda | \eta \text{ off } \Lambda) = \frac{1}{Z'(\Lambda, \eta)} \exp\left(- \sum_{(s,t)} tr \log DF(x^t)_{ss} \right).$$

I came up with this formula in February 1995 (unpublished notes); the same idea was presented independently in [4] and used again in [14].

3.8 Gibbs Phases for Invertible Dynamical Systems

For invertible systems (whether ordinary or CML), symbolic dynamics uses backward-time symbols as well as forward, and the appropriate formulae for Gibbs phases use $\det Df^u$ and $\log Df^u$, where Df^u is the restriction of the derivative of f to the unstable subspace, and some choice of way of comparing volumes on the unstable subspaces of different points has to be made.

For an invertible system, however, it is also possible to ask for the probabilistic behaviour at large negative times. This is given by the backwards Gibbs phases, which use $(Df^s)^{-1}$ instead of Df^u (where Df^s is the restriction of Df to the stable subspace). For systems preserving a volume form, these give the same result, but in general they give different results.

An example where this can be seen by simple simulation (as well as proved) is the torus map

$$x' = 2x + y + \frac{k}{2\pi} \sin 2\pi x$$
$$y' = x + y$$

for $|k| \in (0,1)$. It is uniformly hyperbolic and topologically mixing and has unique forwards and backwards Gibbs phases, but they are different (and singular with respect to each other and Lebesgue measure).

4 Some Challenges

How robust are the examples of indecomposable CML with non-unique phase given here? By Toom's work, Examples 1–4 (and probably 5) are robust to small changes in the slopes provided they remain constant, but presumably one can also make the slopes slightly non-constant and still keep the phenomena, because the resulting Gibbsian random fields on space-time should not be very different from the Markovian ones of the PCA.

Can you make continuous (or smooth) examples of indecomposable CML with non-unique phase? If our examples are very robust then one might be able to make the local map depend continuously on the state of neighbours, rather than just on their symbolic state. Alternatively, to simulate a PCA we chose to use piecewise constant expansion factors, but really we needed this only modulo a coboundary. Does the freedom to choose a coboundary allow to make continuous examples? As a third alternative, might it be possible to make a CML version of Toom's proof for the majority voter PCA? The idea would be to make a continuous CML which is monotone in the usual dynamical systems sense within each local symbol.

Can you explain the "stable chaos" of [11]?

What do Gibbs phases for CML tell us about "typical" orbits? For ordinary dynamical systems, they give the statistics for the orbit of Lebesgue almost every point in the basin of attraction. The problem for spatially extended systems is to make sense of "typical".

What is the most general notion of "nice enough" initial probability distribution which would lead to selection of Gibbs phases?

Can you find examples of non-unique phase without using symbolic dynamics? Unique phase can be proved for some classes of system without symbolic dynamics (e.g. [1] and Keller's lectures), but the methods do not yet suggest how to make examples with non-unique phase. Alternatively, the concept of "equilibrium state" does not require symbolic dynamics, though would exclude nontrivial collective behaviour.

Can you make symplectic examples of reversible CML with non-reversible phase? This would connect more closely to Hamiltonian dynamics and hence to fundamental models of the universe. The symplectic form need not be the standard one.

Can you make a reversible CML exhibiting phases with a second law of thermodynamics, or radiation conditions, rather than just being non-reversible? For example, one might hope to make a form of nontrivial collective behaviour which is not recurrent, or a skew product over the kinetic

Ising model which generates outgoing waves in forward time when the underlying state is + and backward time when it is −. Perhaps a 2D version of Stavskaya's PCA has a phase with an interface zone across which the density of − increases from 0 to its value for the eternal transient phase, and a reversible version could be made?

Appendix: Reversibility of Markov chains and PCA

It is common to regard Markov chains as the epitome of irreversible behaviour: a preferred direction of time seems to be built in because we usually characterise them by their forward transition probabilities, and memory of the initial state is lost as time goes forwards (completely if the chain is irreducible and aperiodic).

Nevertheless, the Markov property does not distinguish between backward and forward time. A Markov chain (e.g. [18]) is a discrete-time stochastic process on a countable state space such that the sequences of states in the past and future of any moment are independent given the state at that moment. From this it follows that the probability of any given finite sequence of states can be factorised as a probability of the initial state multiplied by a sequence of forward transition probabilities, but it could be written equally well as a probability for the final state multiplied by a sequence of backward transition probabilities (or even starting at any intermediate time and working out both ways in time). Note that this is true regardless of whether the Markov chain is stationary (a stochastic process is *stationary* if the probability distribution for realisations is invariant under time-shifts).

A stochastic process $(X_t)_{t\in\mathbb{Z}}$ is said to be *reversible* if the probability distribution for $(X_t, \ldots, X'_t)$ is the same as for $(X_{\tau-t}, \ldots, X_{\tau-t'})$ for all t, t', τ. Note that "reversible" is not the opposite of "irreversible". Most reversible Markov chains exhibit irreversible behaviour, in the sense of loss of dependence of the future state on the state at time 0, but they have identical loss of dependence in reverse time.

I consider it appropriate, even essential, to extend the concept of reversibility of stochastic processes to something which is a closer analogue of the notion in deterministic dynamics. If R is an *involution* of the state space (a map whose square is the identity I), I say a stochastic process $(X_t)_{t\in\mathbb{Z}}$ is *R-reversible* if $(R(X_{\tau-t}), \ldots, R(X_{\tau-t'}))$ has the same distribution as $(X_t, \ldots, X'_t)$ for all t, t', τ. Just as is well known for the standard case of $R = I$, an R-reversible stochastic process is automatically stationary (just choose two different values for τ). Also if π is the probability distribution at time 0 (and hence at any time, by stationarity) for an R-reversible Markov chain with transition probabilities p_{ij} then by taking $t = t' = \tau = 0$ we obtain $\pi_{R(i)} = \pi_i$ for all states i and by taking $t = 0, t' = \tau = 1$ we obtain

$$\pi_i p_{ij} = \pi_{R(j)} p_{R(j)R(i)}$$

for all states i, j (in the standard case $R = I$, the first relation is vacuous and the second is known as "detailed balance"). Furthermore, if a stationary Markov chain satisfies these two conditions then it is easy to prove it is R-reversible (generalising another standard result).

Often, however, under the name "Markov chain" we are presented just with forward transition probabilities p_{ij}, specifying conditional probabilities like $P(i_1, \ldots, i_t|i_0)$ but not $P(i_0)$ nor $P(i_0|i_1)$, for example. This is not a full prescription of a Markov chain, which officially specifies the probabilities of all events in the space of sequences, so to avoid confusion I'll call it a *pre-Markov chain*. It is precisely our situation: we wish to know the possible stochastic processes compatible with the transition probabilities, including arbitrarily far in the past, and decide if there is a unique one or not.

A pre-Markov chain does not define a backwards pre-Markov chain, so is not invertible, yet we'd like to be able to say if a pre-Markov chain is reversible or not. I propose a definition, based on considering the transition probabilities as defining probabilities for two-point boundary value problems rather than initial value problems (as in Sect. 3.3).

To simplify exposition, let's first restrict attention to *autonomous* pre-Markov chains (meaning the transition probabilities do not depend on time). A pre-Markov chain (p_{ij}) defines the probabilities of sequences $i_0, \ldots, i_\tau$ given i_0, i_τ (provided a path of positive probability exists) to be

$$P(i_0, \ldots, i_\tau|\ i_0, i_\tau) = \frac{1}{Z} \prod_{t=0}^{\tau-1} p_{i_t i_{t+1}}$$

with Z a normalisation factor. Given involution R, say the pre-Markov chain is *R-reversible* if for each two-point boundary value problem the probability distribution is the same as for that given by reflecting all sequences in time and in state using R, equivalently,

$$\frac{1}{Z} \prod_{t=0}^{\tau-1} p_{i_t i_{t+1}} = \frac{1}{Z'} \prod_{t=0}^{\tau-1} p_{R(i_{t+1})R(i_t)} \ .$$

This holds if there exists a probability distribution π such that

$$p_{R(j)R(i)} = \pi_i p_{ij}/\pi_j$$

for all i, j, and for a large class of pre-Markov chains this is also a necessary condition (see next four paragraphs). Such a π is an equilibrium measure for the pre-Markov chain ($\sum_i \pi_i p_{ij} = \pi_j$ for all j), so (p, π) defines a Markov chain. If $R = I$ it is reversible, but for $R \neq I$ if $\pi_{R(i)} \neq \pi_i$ for some i then it is not R-reversible. Thus the notion of R-reversibility allows the possibility of making R-reversible pre-Markov chains with non-R-reversible equilibrium measure.

To see that the above sufficient condition is often also necessary, given p for which the graph of allowed transitions is connected and each state can be both

preceded and followed by a cycle (and a hypothesis on the highest common factors of lengths of cycles, to be formulated later), the general solution for non-negative q of

$$\frac{1}{Z}\prod_{t=0}^{\tau-1} p_{i_t i_{t+1}} = \frac{1}{Z'}\prod_{t=0}^{\tau-1} q_{i_t t_{t+1}}$$

for all i_0, i_τ for which there exists an allowed path is

$$q_{ij} = DV_j p_{ij}/V_i$$

for some $D > 0$ and positive function V.

To prove this, let $z_{ij} = q_{ij}/p_{ij}$ for allowed transitions, so the condition is that the product of $z_{i_t i_{t+1}}$ along all paths from i_0 to i_τ in the same time τ is the same. Choose a cycle $i_0, \ldots, i_N = i_0$ and let D be the Nth root of the product of $z_{i_t i_{t+1}}$ around it. Any other cycle gives the same value because if γ_1, γ_2 are two cycles of lengths N_1, N_2, and one can get from γ_1 to γ_2 by a path γ then the two paths $\gamma_1^{N_2}\gamma$ and $\gamma\gamma_2^{N_1}$ connect the same two states in the same time so have the same product of z. Similarly, if γ_1 and γ_2 can be reached from a common point then there exists a preceding cycle γ_3, so γ_1, γ_2 have the same value of D as γ_3 (and similarly if they can lead to a common point). Iterating this procedure for cycles which are linked only by a chain of segments in alternating directions, we show that all cycles have the same value of D.

Now choose a reference state o, let $V_o = 1$ and for each j reachable from o let V_j be the product of the z along a path from o to j divided by D^τ where τ is the length of the path. There might be more than one path from o to j. If they have the length then the product of z is the same, so no ambiguity results. If path γ_1 from o to j is longer than γ_2 by $K > 0$ then j is followed by some path γ_3 ending in a cycle γ. If the length of γ divides K, say N times, then $\gamma_1\gamma_3$ and $\gamma_2\gamma_3\gamma^N$ joins the same points in the same time, so the value of V_j is the same using γ_1 or γ_2. Similarly, if o is preceded by a cycle of length dividing K, or if there are cycles γ_4 preceding o and γ_5 following j whose lengths have highest common factor dividing K, one can construct two paths joining two points in the same time and hence see that V_j is well defined. Similarly if γ_2 contains a state on a cycle with length dividing K, or if both contain states on cycles with lengths whose highest common factor divides K, then we find the same value for V_j. For j that can lead to o, one can define V_j to be D^τ divided by the product (which is non-zero) of z for backward transitions from o, and again argue that this produces no contradiction. Finally, one can alternate forwards and backwards paths to define V_j for all j, because we assumed the graph is connected.

Thus under these assumptions, for an R-reversible pre-Markov chain,

$$p_{R(j)R(i)} = DV_j p_{ij}/V_i$$

for some D and V. Summing over i yields $1 = DV_j \sum_i \frac{p_{ij}}{V_i}$, so $D\sum_i \frac{p_{ij}}{V_i} = \frac{1}{V_j}$. Now assume $Z = \sum_i \frac{1}{V_i} < \infty$ and sum over j, to obtain $DZ = Z$, so $D = 1$.

Then $\pi_i = \frac{1}{Z V_i}$ is a probability distribution such that for all i, j,

$$p_{R(j)R(i)} = \pi_i p_{ij}/\pi_j \ .$$

The considerations of this Appendix extend from Markov chains to PCA. When we talk about PCA, we usually mean a pre-PCA, specifying conditional probabilities for initial value problems, rather than a full stochastic process. This suffices to determine the probabilities for boundary value problems, however, as in Sect. 3.5. Hence given an involution R of the local state space we can define a pre-PCA to be *R-reversible* if the conditional probabilities for all time and state reflected boundary value problems are the same as for the originals. A sufficient condition for this is that the associated Markov random field model be equivalent to one with time-symmetric templates such that the weights for state-time-reflected configurations differ by a (multiplicative) coboundary, i.e. a function of configuration on the template such the product over any bounded subset Λ of space-time given the state outside is independent of the state on Λ (such as V_j/V_i for the case of a pre-Markov chain p_{ij}). One could also allow involutions of the product state space not of product form, but I won't pursue that here.

Now for the kinetic Ising model, which is not autonomous but period-2, we need to modify the definition of reversibility again. I'll say a period-2 Markov chain (meaning the distribution is invariant under shifts by even times) is R-reversible if the distribution of sequences reflected about half-odd integer times and with states reflected by R is unchanged. Similarly, a period-2 pre-Markov chain is R-reversible if the probabilities for all two-point boundary value problems are invariant under reflection in a half-odd integer time and reflection of the states by R. The same goes for period-2 PCA.

For example, the kinetic Ising model is reversible with $R = I$, because an equivalent Markov random field (i.e. after multiplication by a coboundary) is that with templates τ of the form $\{(s,t), (s \pm (1,0), t), (s \pm (0,1), t), (s, t+1), (s \pm (1,0), t+1), (s \pm (0,1), t+1)\}$ for (s,t) even, and weights

$$w_\tau = \frac{\exp \frac{1}{2}(\sigma_s^t + \sigma_s^{t+1})Jm}{\exp Jm + \exp -Jm}$$

if $\sigma_r^{t+1} = \sigma_r^t$ for all $r \in N(s)$ (where $m = \sum_{r \in N(s)} \sigma_r^t$), 0 otherwise, which are invariant under time-reflection. The coboundary consists of factors $\exp \frac{1}{2} J \sigma_s^{t+1} \sigma_r^{t+1}$ and $\exp \frac{1}{2} J \sigma_s^t \sigma_r^t$ for each $r \in N(s)$, each of which cancel with corresponding factors in overlapping templates.

The kinetic Ising PCA is also reversible with respect to interchange of $\pm$.

References

1. J.-B. Bardet, *Limit theorems for coupled analytic maps*, Probab. Th. Rel. Fields **124** (2002) 151–177. 88

2. C. Beck, *Spatio-temporal chaos and vacuum fluctuations of quantized fields* (World Sci, 2002). 76
3. C. Bennett, G. Grinstein, *Role of irreversibility in stabilizing complex and nonergodic behaviour in locally interacting discrete systems*, Phys. Rev. Lett. **55** (1985) 657–660. 67, 71
4. J. Bricmont, A. Kupiainen, *High temperature expansions and dynamical systems*, Commun. Math. Phys. **178** (1996) 703–32. 85, 87
5. L. A. Bunimovich, Ya. G. Sinai, *Space-time chaos in coupled map lattices*, Nonlinearity **1** (1988) 491–516. 65
6. R. Burton R. J. E. Steif, *Non-uniqueness of measures of maximal entropy for subshifts of finite type*, Erg. Th. Dyn. Sys. **14** (1994) 213–35. 72
7. H. Chaté, P. Manneville, *Collective behaviors in spatially extended systems with local interactions and synchronous updating*, Prog. Theor. Phys. **87** (1992) 1–60. 66, 69
8. J. P. Crutchfield, K. Kaneko, *Are attractors relevant to turbulence?*, Phys. Rev. Lett. **60** (1988) 2715–8. 72
9. P. Gacs, *Reliable cellular automata with self-organization*, J. Stat. Phys. **103** (2001) 45–267. 71
10. G. Gielis, R. S. MacKay, *Coupled map lattices with phase transition*, Nonlinearity **13** (2000) 867–888. 66, 67, 71, 72, 74, 85, 86
11. F. Ginelli, R. Livi, A. Politi, *Emergence of chaotic behaviour in linearly stable systems*, J. Phys. A **35** (2002) 499–516. 72, 74, 88
12. L. F. Gray, *A reader's guide to Gacs' "positive rates" paper*, J. Stat. Phys. **103** (2001) 1–44. 71, 72
13. H. Haken, *Cooperative phenomena in systems far from thermal equilibrium and in nonphysical systems*, Rev. Mod. Phys. **47** (1975) 67–121. 65
14. M. Jiang, Ya. B. Pesin, *Equilibrium measures for coupled map lattices: existence, uniqueness and finite dimensional approximations*, Commun. Math. Phys. **193** (1998) 677–711. 66, 87
15. K. Kaneko, *Period-doubling of kink-antikink patterns, quasiperiodicity in antiferro-like structures and spatial intermittency in coupled map lattices*, Prog. Theor. Phys. **72** (1984) 480–486. 66
16. K. Kaneko, *Globally coupled chaos violates the law of large numbers*, Phys. Rev. Lett. **65** (1990) 1391–4. 69
17. A. Katok, B. Hasselblatt, *Introduction to the modern theory of dynamical systems* (Cambridge Univ Press, 1995). 72
18. F. P. Kelly, *Reversibility and stochastic networks* (Wiley, 1979). 71, 89
19. Y. Kuramoto, *Cooperative dynamics of oscillator community*, Prog. Theor. Phys. **79** (1984) 223–240. 66
20. J. L. Lebowitz, C. Maes, E. Speer, *Statistical mechanics of probabilistic cellular automata*, J. Stat. Phys. **59** (1990) 117–70. 70
21. T. M. Liggett, *Interacting particle systems* (Springer, 1985). 74
22. J. Losson, M. C. Mackey, *Thermodynamic properties of coupled map lattices*, in: Stochastic and spatial structures of dynamical systems, eds. S. J. van Strien, S. M. Verduyn Lunel (North Holland, 1996), 41–69. 69
23. R. S. MacKay, Dynamics of networks: features that persist from the uncoupled limit, in "Stochastic and spatial structures of dynamical systems", eds. S. J. van Strien, S. M. Verduyn Lunel (North Holland, 1996), 81–104. 84, 85
24. J. Miller, D. A. Huse, *Macroscopic equilibrium from microscopic irreversibility in a chaotic coupled map lattice*, Phys. Rev. E **48** (1993) 2528–35. 66

25. R. Penrose, *The road to reality* (Random House, 2004) 75
26. Ya. B. Pesin, Ya. G. Sinai, *Space-time chaos in chains of weakly coupled interacting hyperbolic mappings*, Adv Sov Math **3** (1991) 165–198. 84
27. H. Sakaguchi, *Phase transition in coupled Bernoulli maps*, Prog. Theor. Phys. **80** (1988) 7–12. 66, 70, 76
28. Ya. G. Sinai, *Gibbs measures in ergodic theory*, Russ. Math. Surv. **27**:4 (1972) 21–69. 65
29. A. L. Toom, N. B. Vasilyev, O. N. Stavskaya, L. G. Mityushin, G. L. Kurdyumov, S. A. Pirogov, Discrete local Markov systems. In: *Stochastic cellular systems, ergodicity, memory, morphogenesis*, eds R. L. Dobrushin, V. I. Kryukov, A. L. Toom (Manchester Univ Press, 1990), 1–182. 65, 69, 70, 73, 76
30. M. Viana, *Stochastic dynamics of deterministic systems* (Instituto de Matematica Pura e Aplicada, 1997). 65

SRB-Measures for Coupled Map Lattices

E. Järvenpää

Department of Mathematics and Statistics, P.O. Box 35 (MaD), FIN-40014
University of Jyväskylä, Finland
esaj@maths.jyu.fi

This chapter is divided into three parts. I start with a review of the existence results of SRB-measures for coupled map lattices. In the second part I give a brief introduction to the behaviour of Hausdorff dimension under general projections. Finally, I construct a counterexample to the Bricmont–Kupiainen conjecture and discuss its role in the definition of SRB-measures for coupled map lattices.

1 Existence of SRB-measures

A coupled map lattice is a discrete time dynamical system (X, T) consisting of the phase space X and the mapping $T : X \to X$. The phase space $X = \prod_L M$ is the product of a compact (finite dimensional) manifold M over some lattice L equipped with the product topology. The dynamics $T = A \circ F$ is a composition of the uncoupled map $F = \prod_L f$, where $f : M \to M$ is called the local dynamics, and the coupling $A : X \to X$, which ties the independent local systems into one infinite dimensional system. I will mainly concentrate on the situation where M is the unit circle S^1, f is an expanding smooth map, L is the d-dimensional lattice $\mathbb{Z}^d$, and the coupling A is a small perturbation of the identity such that the dependence of the ith coordinate $A(x)_i$ on x_j decreases exponentially fast with respect to the natural distance $|i-j|$ on $\mathbb{Z}^d$.

Coupled map lattices were introduced by Kaneko in 1983. I refer to [40] (see also [11]) for an overview of the numerous results obtained in the 1980's and I will concentrate on the existence results of SRB-measures (for the discussion on the definition of SRB-measure see the end of Sect. 3). The existence of SRB-measures for coupled map lattices was first proved by Bunimovich and Sinai in [12]. They studied a system where the phase space $X = \prod_{\mathbb{Z}}[0,1]$ is the product of unit intervals over the one dimensional lattice $\mathbb{Z}$, the local dynamics $f : [0,1] \to [0,1]$ is an expanding $C^{1+\delta}$ map (that is, f' is Hölder continuous with exponent δ), and A is a special nearest neighbour coupling, meaning that the value of $A(x)_i$ depends only on x_{i-1}, x_i, and x_{i+1}. They

E. Järvenpää: *SRB-Measures for Coupled Map Lattices*, Lect. Notes Phys. **671**, 95–114 (2005)
www.springerlink.com

proved that there exists an ergodic invariant Borel probability measure μ on X such that its projections μ_Λ onto finite dimensional subsystems $X_\Lambda = \prod_\Lambda [0,1]$ ($\Lambda \subset \mathbb{Z}$ is finite) are absolutely continuous with respect to the corresponding Lebesgue measure. The measure μ is also mixing with respect to the dynamics T (which I will call the time dynamics) and the space dynamics induced by the natural left shift $\sigma : \mathbb{Z} \to \mathbb{Z}$, $\sigma(i) = i - 1$. Pesin and Sinai [53] extended this result to the case where the local maps are (genuinely) hyperbolic.

The essential feature of the coupling used in [12, 53] is that it is identity on the boundaries of the space. This guarantees the existence of Markov partitions which enables the use of thermodynamic formalism, that is, one can apply the methods of statistical physics. Keller and Künzle [44] studied a more general coupling and more general local dynamics, namely only piecewise $C^{1+\delta}$, and obtained the existence of an SRB-measure. However, they managed to prove mixing properties only for finite dimensional approximations, not for the whole infinite dimensional system. For corresponding results in the case of hyperbolic local maps, see [60].

In [9] Bricmont and Kupiainen extended the aforementioned results in several ways. First of all, they studied d-dimensional lattice $\mathbb{Z}^d$ for arbitrary $d \in \mathbb{N}$. Secondly, they had an infinite range coupling which was only demanded to be small and to decay exponentially with respect to the distance between lattice sites. Thirdly, they proved good mixing properties, namely, they obtained an exponential decay of correlations for local observables in the infinite dimensional system. The price for this is that they had to consider real analytic expanding maps on the local phase space S^1. In [10] they obtained the same results for $C^{1+\delta}$ maps. The fact that similar methods can be applied if one replaces the unit circle by any compact manifold admitting expanding smooth maps was observed in [31]. Jiang [36] and Jiang and Pesin [37] have proved the corresponding results as Bricmont and Kupiainen in the case where the local dynamics is hyperbolic. Inspired by the works of Just [38, 39], I studied in [30] globally coupled maps and proved the existence of a SRB-measure. Furthermore, I verified that correlations decay like $1/N$ where N is the system size.

A natural question arising from these results is whether one can say something about the spectrum of the Perron–Frobenius operator associated to the infinite dimensional system. This question has been studied for example by Baladi et al. [3], Fischer and Rugh [22], Baladi and Rugh [4], Blank, Keller, and Liverani [7], Schmitt [57], and many others. For more information on this approach see the presentation of Gerhard Keller in this school.

I try to give a rough idea of the proofs of the results in [10] without going too deep into technical details. For readers' convenience I give explicitly the assumptions of [10]. The phase space is $X = \prod_{\mathbb{Z}^d} S^1$. It is practical to use the covering map $x_j \mapsto e^{i2\pi x_j}$ and describe the mappings via their lifts. The local dynamics is given by an orientation preserving expanding $C^{1+\delta}$ map $f : \mathbb{R} \to \mathbb{R}$. Note that this implies the existence of $\lambda > 1$ and $k \in \mathbb{N} \setminus \{0, 1\}$ such that $f'(x) > \lambda$ and $f(x + 1) = f(x) + k$ for all $x \in \mathbb{R}$. One assumes that

$f(0) \in [0,1)$. For all i the ith component $A(x)_i$ of the coupling is assumed to be continuously differentiable with respect to every variable x_j such that for all $x, y \in X$

$$\left| \frac{\partial A(x)_i}{\partial x_j} - \delta_{ij} \right| \leq \varepsilon e^{-\beta|i-j|}$$

and

$$\left| \frac{\partial A(x)_i}{\partial x_j} - \frac{\partial A(y)_i}{\partial x_j} \right| \leq \varepsilon \sum_{k \in \mathbb{Z}^d} e^{-\beta(|i-j|+|i-k|)} |x_k - y_k|^\delta$$

for some $\varepsilon, \beta, \delta > 0$. Here δ_{ij} is the Kronecker's delta. Furthermore, the coupling is assumed to be space translation invariant, that is, $A = \tau_i \circ A \circ \tau_i^{-1}$ for all $i \in \mathbb{Z}^d$. Here $(\tau_i(x))_j = x_{j-i}$.

Example 1.1. An example of a coupling satisfying the above assumptions is

$$A(x)_i = x_i + \frac{\varepsilon}{(2\pi)^2} \sum_{k \in \mathbb{Z}^d} 2^{-|i-k|} \cos(2\pi(x_i - x_k))$$

with $\beta = \log 2/2$.

Remark 1.1. In this chapter I always assume that measures are Borel regular (for the definition see [49, Definition 1.5]). Since Bricmont and Kupiainen defined a regular measure meaning something different (see Definition 1.2) I do not explicitly mention Borel regularity to avoid confusion.

Definition 1.1. *A measure μ is T-invariant if $T_*\mu = \mu$ where the image measure is defined by $T_*\mu(B) = \mu(T^{-1}(B))$ for all Borel sets B. A T-invariant probability measure μ on X is a SRB-measure if $\mu_\Lambda \ll \mathcal{L}^\Lambda$ for all finite $\Lambda \subset \mathbb{Z}^d$ where $\mathcal{L}$ is the Lebesgue measure on X. The notation $\mu \ll \nu$ means that measure μ is absolutely continuous with respect to ν. If $\mu \ll \nu$ then the Radon–Nikodym derivative $d\mu/d\nu$ is called the density of μ with respect to ν.*

For the differentiation theory of measures see [49, Sect. 2].

Definition 1.2. *Let $\Lambda \subset \mathbb{Z}^d$ be finite. The space $C^\delta(X_\Lambda)$ consists of continuous functions h on X_Λ such that*

$$|h(x) - h(y)| \leq C \sum_{i \in \Lambda} |x_i - y_i|^\delta.$$

We equip $C^\delta(X_\Lambda)$ with the norm

$$\|h\|_\delta = \|h\|_\infty + \sup_{x,y \in X_\Lambda} \frac{|h(x) - h(y)|}{\sum_{i \in \Lambda} |x_i - y_i|^\delta} \ .$$

A probability measure μ on X is regular if $\mu_\Lambda \ll \mathcal{L}^\Lambda$, its density h_Λ is continuous and strictly positive, and $\log h_\Lambda \in C^\delta(X_\Lambda)$ for all finite $\Lambda \subset \mathbb{Z}^d$.

In [10] the following theorem was proved.

Theorem 1.1. *Let T satisfy the assumptions given above. There exists $\varepsilon_0 > 0$ such that for all $\varepsilon < \varepsilon_0$ the map T has a SRB-measure μ. Furthermore, μ is invariant and exponentially mixing under the space-time translations generated by τ_i and T: there exists $m > 0$ and $C < \infty$ such that for all finite $B, D \subset \mathbb{Z}^d$ and for all $g \in L^\infty(X_B)$ and $h \in C^\delta(X_D)$ we have*

$$\left| \int g \circ T^n h d\mu - \int g d\mu \int h d\mu \right| \leq C \ \|g\|_\infty \ \|h\|_\delta \ e^{-m(n+d(B,D))}$$

where $d(B, D) = \min\{|i - j| \mid i \in B, j \in D\}$ and C depends on the diameters of B and D. Finally, for all regular measures ν the sequence $(T_*)^n \nu$ tends to μ in the weak* topology.

There are three main tools in the proof of Theorem 1.1: Perron–Frobenius operator, equilibrium states associated to interactions in lattice models of statistical physics, and polymer expansions.

Definition 1.3. *Let M be a compact manifold, $\mathcal{L}$ the Lebesgue measure on M, and $T : M \to M$ a non-singular mapping meaning that $\mathcal{L}(B) = 0$ if and only if $\mathcal{L}(T^{-1}(B)) = 0$ for Borel sets B. The Perron–Frobenius operator $P : L^1(M) \to L^1(M)$ associated to T is defined by the equation*

$$\int g \circ T \ h \, d\mathcal{L} = \int g P(h) \, d\mathcal{L} \tag{1}$$

for all $g \in L^\infty(M)$ and $h \in L^1(M)$.

Observe that if $\mu \ll \mathcal{L}$ then also $T_*\mu \ll \mathcal{L}$ by the non-singularity of T. Thus if h is the Radon–Nikodym derivative of μ with respect to $\mathcal{L}$ then $P(h)$ is the Radon–Nikodym derivative of $T_*\mu$ with respect to $\mathcal{L}$. (In the sequel the Radon–Nikodym derivative is called simply the density.) Therefore, P may be viewed as the restriction of T_* to measures which are absolutely continuous with respect to $\mathcal{L}$. From this it is obvious that if h is a fixed point of P, that is, it is an eigenstate with eigenvalue 1, then $h\mathcal{L}$ is an invariant measure for T. It is a well-known result that if T is an expanding $C^{1+\delta}$ map on a compact manifold M then the spectrum of P, restricted to the space of functions of bounded variation, consists of a simple eigenvalue 1 and the rest of the spectrum is inside a disc of radius γ for some $\gamma < 1$. This implies that any smooth initial density will converge under the iteration of P to the density of the unique invariant measure μ which is absolutely continuous with respect to $\mathcal{L}$. In fact, the SRB-measure μ is equivalent with $\mathcal{L}$. These results are due to Sinai [59], Ruelle [54], and Bowen [8]. For more information on these questions see [1, 2]. An easy-to-access exposition of interval maps is [13].

I will need the following explicit form of the Perron–Frobenius operator which is valid in the expanding orientation preserving case:

$$P(h)(x) = \sum_{y:T(y)=x} \frac{h(y)}{\det DT(y)} \cdot \tag{2}$$

This formula is easily verified via the change of variables.

The idea of the proof of Theorem 1.1 is that one first takes a finite $\Lambda \subset \mathbb{Z}^d$, fixes some boundary condition x_{Λ^c} on X_{Λ^c} (where $\Lambda^c = \mathbb{Z}^d \setminus \Lambda$), and obtains in this way an expanding map on the finite dimensional manifold X_Λ. More precisely, $T_\Lambda : X_\Lambda \to X_\Lambda$ is defined by $T_\Lambda = A_\Lambda \circ \prod_\Lambda f$, where $A_\Lambda(x)_i = A(x_\Lambda, x_{\Lambda^c})_i$ for all $i \in \Lambda$ and the boundary condition x_{Λ^c} does not evolve in time. Then one iterates the constant function 1 with the Perron–Frobenius operator P_Λ associated to this finite dimensional approximation. By the aforementioned results concerning finite dimensional expanding maps the sequence $P_\Lambda^N(1)$ will tend to the density of the unique SRB-measure μ_Λ of the finite dimensional system. Finally, letting Λ tend to $\mathbb{Z}^d$ along some increasing sequence, one shows that the sequence μ_Λ converges to some measure μ which is independent of the boundary condition x_{Λ^c}. This limiting measure is the SRB-measure of Theorem 1.1.

As a result of the above method one obtains the following lattice gas model. In addition to the lattice $\mathbb{Z}^d$ each iterate of the Perron–Frobenius operator will give a new copy of $\mathbb{Z}^d$ leading to the lattice $L = \mathbb{Z}^d \times \mathbb{N}$. As a result we have a phase space $\Omega = \prod_L M_{(i,j)}$, where $M_{(i,0)} = S^1$ for all $i \in \mathbb{Z}^d$ and $M_{(i,j)} = \{0, \dots, k-1\}$ for all $i \in \mathbb{Z}^d$ and $j \in \mathbb{N} \setminus \{0\}$. Here k is the number of preimages of any point for the map $f : S^1 \to S^1$. Denoting by f also the lift of f, one sees that the different preimages of a point $x \in [0,1)$ are of the form $f^{-1}(x+s)$ where $s = 0, 1, \dots, k-1$. Let $\psi_s(x)_i = f^{-1}(A_\Lambda^{-1}(x)_i + s_i)$ where $s \in \{0, \dots, k-1\}^\Lambda$. Then by (2)

$$P_\Lambda(h)(x) = \det DA_\Lambda^{-1}(x) \sum_{s \in \{0,\dots,k-1\}^\Lambda} \frac{h(\psi_s(x))}{\prod_{i\in\Lambda} f'(\psi_s(x)_i)}$$

and further

$$P_\Lambda^N(1)(x) = \sum_{s^1,\dots,s^N} \prod_{t=1}^{N} \Big(\det DA_\Lambda^{-1}(\psi_{s^{t-1}} \circ \cdots \circ \psi_{s^1}(x)) \prod_{i\in\Lambda} (f'(\psi_{s^t} \circ \cdots \circ \psi_{s^1}(x)_i))^{-1} \Big) \cdot$$

One begins the proof by writing

$$P_\Lambda^N(1) = e^{-H} \tag{3}$$

where H is a real valued function on $\Omega_{\Lambda\times\{0,\dots,N\}}$. In the terminology of statistical physics the function H is called Hamiltonian. In order to apply the methods of statistical physics one still has to localize the Hamiltonian, meaning that one needs to find an interaction Φ such that

$$H = \sum_{Y \subset \Lambda \times \{0,\dots,N\}} \Phi_Y$$

where Φ_Y is a function on Ω_Y. One proves that the interaction Φ_Y, which depends on Λ, N, and the chosen boundary condition x_{Λ^c}, will tend to an interaction Φ_Y, which is independent of Λ, N, and the boundary condition x_{Λ^c}, as N tends to infinity and Λ tends to $\mathbb{Z}^d$. The results of statistical mechanics guarantee that, provided the limiting interaction Φ has some nice properties, there will be a unique equilibrium state $\tilde{\mu}$ associated to it. This equilibrium state is a measure on Ω and it will be exponentially mixing with respect to translations in L. Note that the translation in $\mathbb{N}$ direction corresponds to the time dynamics of the original coupled map lattice. Thus the projection of $\tilde{\mu}$ onto $\Omega_{\mathbb{Z}^d \times \{0\}}$ will be the SRB-measure of Theorem 1.1.

An excellent introduction and a handbook of lattice gas models of statistical physics is [58]. A more compressed but still complete exposition can be found in [55]. I refer to these books for more information on equilibrium and Gibbs states and I briefly explain some relevant results concerning interactions. In the case where the underlying lattice is at least two dimensional, it is important to distinguish between several norms used in the space of interactions. For simplicity I assume that the interactions are translation invariant and 0 is a fixed point in the lattice. If

$$\|\Phi\|_0 = \sum_{Y \ni 0} |\Phi_Y|$$

is finite, then there exists an equilibrium state (see [58, Corollary III.2.9]). Assuming that

$$\|\Phi\|_1 = \sum_{Y \ni 0} |Y||\Phi_Y|$$

is small, the equilibrium state is unique (see [14, 15]). If for some $\gamma > 0$

$$\|\Phi\|_2 = \sum_{Y \ni 0} e^{\gamma d(Y)} |\Phi_Y|$$

is small, then the equilibrium state is unique and the correlations decay exponentially (see [23]). Finally, supposing that for some $\gamma > 0$

$$\|\Phi\|_3 = \sum_{Y \ni 0} e^{\gamma |Y|} |\Phi_Y|$$

is small, the equilibrium state is unique and the correlation functions are analytic (see [29]). Here $|Y|$ is the number of elements in Y and $d(Y)$ is the diameter of Y.

The interaction rising from the proof of Theorem 1.1 does not fall into any of the above classes. In fact, it is of the form $\Phi^0 + \Phi^1$ where Φ^0 is a finite range one dimensional interaction and $\|\Phi^1\|_2$ is small. In [9] the analyticity

assumptions guarantee that $\|\Phi^1\|_3$ is small. Since the behaviour of one dimensional lattice models is radically different from higher dimensional ones, this kind of interaction will imply the nice properties stated in Theorem 1.1.

How does one localize a Hamiltonian? One method (also used by Bricmont and Kupiainen) is telescoping. Namely, let $H = H(x_1, \ldots, x_n)$ and 0 be some fixed value which every coordinate may attain. Then

$$\begin{aligned} H(x_1, \ldots, x_n) =& H(x_1, \ldots, x_n) - H(x_1, \ldots, x_{n-1}, 0) \\ &+ H(x_1, \ldots, x_{n-1}, 0) - H(x_1, \ldots, x_{n-2}, 0, 0) \\ &+ \cdots + H(x_1, 0, \ldots, 0) - H(0, \ldots, 0) + H(0, \ldots, 0) \ . \end{aligned}$$

Now one can define

$$\Phi_{\{1,\ldots,k\}}(x_1, \ldots, x_k) = H(x_1, \ldots, x_k, 0, \ldots, 0) - H(x_1, \ldots, x_{k-1}, 0, \ldots, 0)$$

and set $\Phi_Y \equiv 0$ if $Y \neq \{1, \ldots, k\}$ for all $k = 1, \ldots, n$. To use this method, some numbering of points in $\Lambda \times \{0, \ldots, N\}$ has to be fixed. The ε in the coupling guarantees that if the set Y contains points with different space coordinates, then there will be ε in the interaction. The exponential decay of the coupling, in turn, is needed to make the norm $\|\cdot\|_2$ small. Finally, one has to decompose the Hamiltonian also in the time direction. Note that in time direction one has a sequence of ψ's (see (3)). Each ψ is a contraction with coefficient $\lambda' < 1$ since f^{-1} is a contraction and the coupling is small. Although λ' is not necessarily small, $(\lambda')^M$ is small for M large. So one has to divide the terms into two groups depending on the number of ψ's they contain. This is the reason why one has the Φ^0-part in the interaction which is not small.

Since the interaction does not fall into any of the well-known classes of interactions described above, the claims of Theorem 1.1 have to be proved explicitly. It is clear that if there is some correlation between observables supported on B and D there must be interactions connecting the sets B and D. There are two possibilities: either there is one term Φ_Y connecting these sets, or there is a chain of terms $\Phi_{Y_1}, \ldots, \Phi_{Y_n}$ starting from B and ending in D. In the first case, the smallness of correlation is due to the exponential decay of the interaction with respect to the diameter of the set. In the second case, there must be many terms and each of them contains one ε. To make this argument rigorous one has to have a good way of indexing all different possibilities. The polymer expansion is a technical tool which is developed for these purposes. It enables one to write an explicit formula for the correlation functions and expectations of local observables, that is, functions depending only on finitely many coordinates. From this formula one finds an upper bound for the rate of mixing. I refer to [9, 10] for more details. An excellent introduction to polymer expansions is [58, Sect. V.7].

In [9] Bricmont and Kupiainen made a conjecture that there is only one SRB-measure for the systems they considered. Note that according to Theorem 1.1 this conjecture is true in the class of regular measures (see Definition 1.2).

2 Projection Results

In this section I give an overview of some results in the field of geometric measure theory which seem to be totally unrelated to the results of the previous section. However, it appears that they are closely connected to the solution of Bricmont–Kupiainen conjecture.

I start with a simple example.

Example 2.1. Consider a line segment in $\mathbb{R}^3$. If one projects it onto a two dimensional plane, one obtains a line segment, unless the plane happens to be perpendicular to the original line segment in which case one obtains a point. Hence this one dimensional set is projected generically onto a one dimensional set. If, instead of line segment, one projects a cube the result is two dimensional. Thus the projection of the three dimensional cube is two dimensional.

Amazingly these two simple examples illustrate the general behaviour of Hausdorff dimensions of projections as the following theorem states:

Theorem 2.1. *Let $0 < m < n$ be integers and $B \subset \mathbb{R}^n$. Then*

$$\dim_{\mathrm{H}}(P_V(B)) = \min\{\dim_{\mathrm{H}}(B), m\}$$

for $\gamma_{n,m}$-almost all $V \in G(n,m)$. Here $G(n,m)$ is the Grasmann manifold of m-planes in $\mathbb{R}^n$, $\gamma_{n,m}$ is the Haar measure on $G(n,m)$, P_V is the orthogonal projection onto V, and $\dim_{\mathrm{H}}(B)$ is the Hausdorff dimension of B.

This theorem was first proven by Marstrand [46] in the plane, then by Kaufman [41] in the plane using different methods, and finally by Mattila [47] in full generality. The corresponding result is true also for measures. Recall that the definition of the Hausdorff dimension of a measure is as follows:

Definition 2.1. *The Hausdorff dimension of a finite measure μ on $\mathbb{R}^n$ is*

$$\begin{aligned}\dim_{\mathrm{H}}(\mu) &= \inf\{\dim_{\mathrm{H}}(B) \mid B \text{ is a Borel set with } \mu(B) > 0\} \\ &= \mu\text{-}\operatorname*{ess\,inf}_{x\in\mathbb{R}^n} \underline{\dim}_{\mathrm{loc}}\,\mu(x)\end{aligned}$$

where

$$\underline{\dim}_{\mathrm{loc}}\,\mu(x) = \liminf_{r\to 0} \frac{\log \mu(B(x,r))}{\log r}$$

is the lower local dimension of μ at x. Here $B(x,r)$ is the closed ball centred at x with radius r.

The analogue of Theorem 2.1 is valid for measures:

Theorem 2.2. *Let $0 < m < n$ be integers and let μ be a finite measure on $\mathbb{R}^n$. Then for $\gamma_{n,m}$-almost all $V \in G(n,m)$*

$$\begin{aligned}\dim_{\mathrm{H}}(P_{V*}\mu) = \dim_{\mathrm{H}}(\mu), &\text{ if } \dim_{\mathrm{H}}(\mu) \le m \\ P_{V*}\mu \ll \mathcal{L}^m, &\text{ if } \dim_{\mathrm{H}}(\mu) > m\,.\end{aligned}$$

Theorem 2.2 is more or less explicitly stated in [20, 24, 41, 48]. Theorems 2.1 and 2.2 may be summed up by saying that Hausdorff dimension is preserved under typical projections. Note that the set of exceptional directions may be quite large although it has zero measure (see [16, 42, 52]).

There are several extensions of these theorems. Instead of Hausdorff dimension, one may consider different dimensions like packing dimension, box counting dimension, multifractal q-dimensions etc. (see [18, 19, 20, 21, 25, 32, 35]). Note that the choice of dimension effects also on the results obtained. One may also replace typical projections by a prevalent set of C^1-maps (see [25, 32, 56]). Hunt and Kaloshin [26] showed that the corresponding results are not true in infinite dimensional situations. I refer to the excellent survey of Mattila [50] for more information.

To give some flavour of the aforementioned results, I make a simple calculation which contains some essential features related to projections. One crucial ingredient in this context is the s-energy of a measure which I now define.

Definition 2.2. *Let μ be a finite measure on a metric space (X, d) and $s > 0$. The s-energy of μ is*

$$I_s(\mu) = \iint d(x,y)^{-s} d\mu(x) d\mu(y) \, .$$

The finiteness of s-energy is closely related to the Hausdorff dimension of a measure. In fact, if $I_s(\mu) < \infty$ then $\dim_{\mathrm{H}}(\mu) \geq s$. On the other hand, assuming that $s < \dim_{\mathrm{H}}(\mu)$, the measure μ has a restriction with finite s-energy (see [49, pp. 109–110]). This indicates the relevance of the energies of projected measures when calculating their Hausdorff dimensions. To begin with I derive a useful formula for integrals according to which for any measure μ on a separable metric space X

$$\int f(x) d\mu(x) = \int_0^\infty \mu(\{x \in X \mid f(x) \geq t\}) \; d\mathcal{L}^1(t) \tag{4}$$

for all non-negative Borel functions f (see [49, Theorem 1.15]). Indeed, letting $A = \{(x,t) \in X \times \mathbb{R} \mid f(x) \geq t\}$ we have

$$\begin{aligned}
&\int_0^\infty \mu(\{x \in X \mid f(x) \geq t\}) \; d\mathcal{L}^1(t) = \int_0^\infty \mu(\{x \in X \mid (x,t) \in A\}) \; d\mathcal{L}^1(t) \\
&= \int \mathcal{L}^1(\{t \in [0,\infty) \mid (x,t) \in A\}) \; d\mu(x) = \int \mathcal{L}^1([0, f(x)]) \; d\mu(x) \\
&= \int f(x) \; d\mu(x).
\end{aligned}$$

Let μ be a measure on $\mathbb{R}^n$. The starting point of the energy calculations of projections of μ is [49, Lemma 3.11] according to which there exists a constant c such that

$$\gamma_{n,m}(\{V \in G(n,m) \mid |P_V(x)| \le \delta\}) \le c\delta^m |x|^{-m}. \tag{5}$$

Combining this with (4), we obtain for all $0 < s < m$ and $x \in \mathbb{R}^n \setminus \{0\}$ the following result (see [49, Corollary 3.12]):

$$\begin{aligned}
&\int |P_V(x)|^{-s} d\gamma_{n,m}(V) = \int_0^\infty \gamma_{n,m}(\{V \in G(n,m) \mid |P_V(x)|^{-s} \ge t\})\, d\mathcal{L}^1(t) \\
&= \int_0^\infty \gamma_{n,m}(\{V \in G(n,m) \mid |P_V(x)| \le t^{-\frac{1}{s}}\})\, d\mathcal{L}^1(t) \\
&\le \int_0^{|x|^{-s}} d\mathcal{L}^1(t) + \int_{|x|^{-s}}^\infty \gamma_{n,m}(\{V \in G(n,m) \mid |P_V(x)| \le t^{-\frac{1}{s}}\})\, d\mathcal{L}^1(t) \\
&\le |x|^{-s} + c|x|^{-m} \int_{|x|^{-s}}^\infty t^{-\frac{m}{s}}\, d\mathcal{L}^1(t) \\
&= \left(1 + \frac{cs}{m-s}\right) |x|^{-s}\,.
\end{aligned} \tag{6}$$

Using (6), we find a constant C such that

$$\begin{aligned}
\int I_s(P_{V*}\mu) d\gamma_{n,m}(V) &= \iiint |x-y|^{-s} dP_{V*}\mu(x) dP_{V*}\mu(y) d\gamma_{n,m}(V) \\
&= \iiint |P_V(x-y)|^{-s} d\mu(x) d\mu(y) d\gamma_{n,m}(V) \\
&= \iiint |P_V(x-y)|^{-s} d\gamma_{n,m}(V) d\mu(x) d\mu(y) \le C I_s(\mu).
\end{aligned} \tag{7}$$

According to (7), if $I_s(\mu) < \infty$ then $I_s(P_{V*}\mu) < \infty$ for $\gamma_{n,m}$-almost all V. This is one of the key observations of the proofs of Theorems 2.1 and 2.2. It is clear that this method will give results only for almost all projections. However, as I mentioned earlier, "almost all"-results are the best possible ones in this context.

A very important extension of the projection results is due to Peres and Schlag [52]. They replace projections by a general parametrized transversal family of mappings from a compact metric space to $\mathbb{R}^m$. Intuitively, transversality means that when the parameter is also changed the mapping is changed fast enough. Since I will use [52, Theorem 7.3] in Sect. 3, I state it explicitly but not in the full generality.

Definition 2.3. *Let (X,d) be a compact metric space, $Q \subset \mathbb{R}^n$ an open connected set, and $\Pi : Q \times X \to \mathbb{R}^m$ a continuous map with $n \ge m$. Define for all $x \ne y \in X$*

$$\Phi_{x,y}(\lambda) = \frac{\Pi(\lambda,x) - \Pi(\lambda,y)}{d(x,y)}\,.$$

The mapping Π is regular if for any multi-index $\eta = (\eta_1, \dots, \eta_n) \in \mathbb{N}^n$ there exists a constant C_η such that

$$|\partial^\eta \Pi(\lambda, x)| \le C_\eta \text{ and } |\partial^\eta \Phi_{x,y}(\lambda)| \le C_\eta$$

for all $\lambda \in Q$ *and for all* $x \neq y \in X$. *Here* $\partial^\eta = \frac{\partial^{|\eta|}}{(\partial\lambda_1)^{\eta_1}\ldots(\partial\lambda_n)^{\eta_n}}$ *and* $|\eta| = \sum_{i=1}^n \eta_i$.

Definition 2.4. *A regular mapping* $\Pi : Q \times X \to \mathbb{R}^m$ *is transversal if there exists a constant* C *such that for all* $\lambda \in Q$ *and for all* $x \neq y \in X$ *the condition* $|\Phi_{x,y}(\lambda)| \le C$ *implies*

$$\det(D\Phi_{x,y}(\lambda)(D\Phi_{x,y}(\lambda))^T) \ge C^2.$$

Here the derivative with respect to λ *is denoted by* D *and* A^T *is the transpose of a matrix* A.

The following theorem from [52] gives a relation between Sobolev norms of images of measures under transversal family of regular mappings and energies of original measures.

Theorem 2.3. *Let* $\Pi : Q \times X \to \mathbb{R}^m$ *be transversal and let* μ *be a finite measure on* X *such that* $I_s(\mu) < \infty$ *for some* $s > 0$. *Then there exists a constant* C_γ *such that*

$$\int_Q \|\Pi_*\mu\|_{2,\gamma}^2 \, d\mathcal{L}^n(\lambda) \le C_\gamma I_s(\mu)$$

provided that $m + 2\gamma \le s$. *Here* $\|\cdot\|_{2,\gamma}$ *is the Sobolev norm, that is,*

$$\|\nu\|_{2,\gamma}^2 = \int_{\mathbb{R}^m} |\hat{\nu}(\xi)|^2 |\xi|^{2\gamma} d\mathcal{L}^m(\xi)$$

for any finite compactly supported measure on $\mathbb{R}^m$, *where*

$$\hat{\nu}(\xi) = \int_{\mathbb{R}^m} e^{-i\xi\cdot x} d\nu(x)$$

is the Fourier transform of ν.

Proof. See [52, Theorem 7.3].

Remark 2.1. Let ν be a finite compactly supported measure on $\mathbb{R}^n$. If $\|\nu\|_{2,0} < \infty$ then ν is absolutely continuous with respect to the Lebesgue measure $\mathcal{L}^n$ and its Radon-Nikodym derivative is L^2-integrable (see [33, Remark 2.6]).

I stated only the part of [52, Theorem 7.3] concerning measures whose dimensions are larger than m since I need only that part in Sect. 3. There is also the part which says that the dimension is preserved if it is less than m. I say briefly something about the assumptions of Theorem 2.3. It is quite obvious that some kind of smoothness is needed and almost as obvious that the regularity assumption I formulated in Definition 2.3 is not the optimal

one. The role of transversality may be explained as follows. Recall from the results of orthogonal projections that there may always be some exceptional directions. Since Theorem 2.3 deals with a parametrized family of maps one could smoothly parametrize one exceptional projection with as many parameters as one likes. Thus there must be some condition guaranteeing that if the parameter is changed also the mapping is changed. In the proof there will be a calculation like (7) in one form or the other where one needs an analogue of (5). Transversality, in turn, is needed for this.

3 Counterexample to Bricmont–Kupiainen Conjecture

In this section I describe how one can construct a counterexample to the Bricmont–Kupiainen conjecture (see [33]). This counterexample is based on the results of Sect. 2.

Consider the standard $\frac{1}{3}$-Cantor set K on the unit interval and let $\mathcal{H}^s|_K$ be the restriction of the s-dimensional Hausdorff measure to K where $s = \log 2/\log 3$ is the Hausdorff dimension of K. Then $\mathcal{H}^s|_K$ is an ergodic invariant measure for the map $f : [0,1] \to [0,1]$, $f(x) = 3x \mod 1$. Since the packing and Hausdorff dimensions of K coincide, the Hausdorff dimension of the n-fold product K^n is equal to $n \log 2/\log 3$ (see [49, Theorem 8.10]). Thus for any m one can find N such that for all $n > N$ the dimension of K^n is larger than m. This is true also for the product measure $(\mathcal{H}^s|_K)^n$. According to Theorem 2.2, the projection of $(\mathcal{H}^s|_K)^n$ is absolutely continuous with respect to $\mathcal{L}^m$ for typical m-planes. However, the Bricmont–Kupiainen conjecture deals with specific projections, namely, those determined by coordinate planes. And clearly in this example the coordinate planes are not typical. The idea of our example is that a small coupling will perturb the uncoupled map slightly such that the coordinate planes become typical ones. Actually we do not prove quite this but instead we find a parametrized family of conjugating maps such that almost all conjugates of the uncoupled map have infinitely many SRB-measures.

Let $X = \prod_{\mathbb{Z}^d} S^1$, $\mathcal{K} = \prod_{\mathbb{Z}^d} K$, and $\mu = \prod_{\mathbb{Z}^d} \mathcal{H}^s|_K$. Since $\mathcal{H}^s(K) = 1$ (see [17, Theorem 1.14]), μ is a probability measure. I denote by μ_Λ the projection of μ onto X_Λ. As in Sect. 1, I will use the same symbol for maps (like $f : S^1 \to S^1$, $f(z) = z^3$) and their lifts (like $f : \mathbb{R} \to \mathbb{R}$, $f(x) = 3x$). Consider $A_\epsilon : X \to X$,

$$A_\varepsilon(x)_i = x_i + \sum_{l \in \mathbb{Z}^d} \varepsilon_{il} 2^{-|i-l|} g(x_l)$$

where for some $\varepsilon_0 > 0$ one has $\varepsilon_{ij} \in (-\varepsilon_0, \varepsilon_0)$ for all $i, j \in \mathbb{Z}^d$, and $g : \mathbb{R} \to \mathbb{R}$ is smooth and 1-periodic. For convenience assume that $|g(x)| < 1$ for all $x \in \mathbb{R}$. It is not difficult to see that A_ε is invertible provided ε_0 is small enough (depending on $|g'|$). Namely, local invertibility follows from the

implicit function theorem, and global invertibility from the fact that A_ε is a lift of a map on X. Set $E = \prod_{\mathbb{Z}^d \times \mathbb{Z}^d}(-\varepsilon_0, \varepsilon_0)$ and $\mathcal{L} = \prod_{\mathbb{Z}^d \times \mathbb{Z}^d} \rho$ where ρ is the normalized restriction of the Lebesgue measure to $(-\varepsilon_0, \varepsilon_0)$. I use the abbreviations $E_{\Lambda\times\tilde\Lambda}$ and $\mathcal{L}^{\Lambda\times\tilde\Lambda}$ for the natural restrictions. For $\Lambda \subset \tilde\Lambda$, I denote the natural projection from $X_{\tilde\Lambda}$ onto X_Λ by $\pi_{\tilde\Lambda,\Lambda}$. Let $T_\varepsilon = A_\varepsilon \circ \prod_{\mathbb{Z}^d} f \circ A_\varepsilon^{-1}$. Now $A_{\varepsilon *}\mu$ is clearly T_ε-invariant. In [33] we proved the following theorem.

Theorem 3.1. *For $\mathcal{L}$-almost all $\varepsilon \in E$ the map T_ε has infinitely many SRB-measures.*

The proof is divided into two steps. First we fix finite $\Lambda \subset \tilde\Lambda \subset \mathbb{Z}^d$ such that $|\tilde\Lambda|s > |\Lambda|$ and consider the restriction of A to $\tilde\Lambda$ with open boundary condition, that is,

$$A_{\varepsilon,\tilde\Lambda}(x)_i = x_i + \sum_{l\in\tilde\Lambda} \varepsilon_{il} 2^{-|i-l|} g(x_l) \; .$$

Using Theorem 2.3, we show that for $\mathcal{L}^{\Lambda\times\tilde\Lambda}$-almost all $\varepsilon \in E_{\Lambda\times\tilde\Lambda}$ the projection $(\pi_{\tilde\Lambda,\Lambda} \circ A_{\varepsilon,\tilde\Lambda})_* \mu_{\tilde\Lambda}$ is absolutely continuous. For this we have to restrict the map g by demanding that $|g'| > b > 0$ in the set K. (Note that by 1-periodicity there must be points where $g' = 0$.) This guarantees that the transversality condition is satisfied. The final step is to let $\tilde\Lambda$ tend to $\mathbb{Z}^d$. This is a technical calculation and I refer to [33] for more details. Intuitively this limiting process should not cause any problems. Indeed, according to Theorem 2.3, the larger the dimension of the measure, the smoother its projection is. Thus tilting the measure also in the complement of $\tilde\Lambda$ should make the situation better. However, one should keep in mind the result of Hunt and Kaloshin [26] according to which the Hausdorff dimension may decrease under projections from infinite dimensional space onto finite dimensional subspaces. Although it is possible that everything fails at the infinite limit, our setup indicates that this is not very likely and the smoothness argument given by Theorem 2.3 is stronger. The technical calculation in [33, Proposition 3.3] shows that this is indeed the case.

Finally, the above consideration implies that there are at least two SRB-measures, namely $A_{\varepsilon *}\mu$ and the SRB-measure constructed by Bricmont and Kupiainen. However, instead of taking the standard Cantor set where one deletes the middle third, one may delete the first or the last third. At every coordinate direction one may choose one of these three sets. Clearly it is possible to choose g such that $|g'| > b > 0$ on all these three sets. In this way one obtains infinitely many SRB-measures.

Now I explain some details of the proof to illustrate what is going on. To use Theorem 2.3, let $\Pi : E_{\Lambda\times\tilde\Lambda} \times Y_{\tilde\Lambda} \to X_\Lambda$, $\Pi(\varepsilon, x) = \pi_{\tilde\Lambda,\Lambda} \circ A_{\varepsilon,\tilde\Lambda}(x)$. Here $Y_{\tilde\Lambda} = \prod_{\tilde\Lambda}[-t_0, t_0]$ for some $\frac{1}{9} < t_0 < \frac{1}{6}$. This is just a technical detail to guarantee that $|g(t) - g(t')| \geq b|t - t'|$. Now Π is clearly regular so it is enough to study the validity of the transversality assumption. We equip $Y_{\tilde\Lambda}$ with the metric

$$d(x,y)^2 = \sum_{l\in\tilde{\Lambda}} 2^{-2|i_0-l|}|x_l - y_l|^2$$

where $i_0 \in \Lambda$ is some fixed point. Now

$$\Phi_{x,y}(\varepsilon)_i = \frac{\Pi(\varepsilon,x) - \Pi(\varepsilon,y)}{d(x,y)} = \frac{x_i - y_i + \sum_{l\in\tilde{\Lambda}} \varepsilon_{il} 2^{-|i-l|}(g(x_l) - g(y_l))}{d(x,y)}$$

for all $i \in \Lambda$. Fix $i \in \Lambda$, $k = (k_1, k_2) \in \Lambda \times \tilde{\Lambda}$, and $x \neq y \in Y_{\tilde{\Lambda}}$. Then

$$D\Phi_{x,y}(\varepsilon)_{i,k} = \frac{\delta_{i,k_1} 2^{-|i-k_2|}(g(x_{k_2}) - g(y_{k_2}))}{d(x,y)} \ .$$

Thus for $i, j \in \Lambda$

$$\begin{aligned}(D\Phi_{x,y}(\epsilon)D\Phi_{x,y}(\epsilon)^T)_{i,j} =& \frac{\delta_{i,j}}{d(x,y)^2} \sum_{l\in\tilde{\Lambda}} 2^{-|i-l|-|j-l|}(g(x_l) - g(y_l))^2 \\ &\geq \delta_{i,j} b^2 2^{-|i-i_0|-|j-i_0|}.\end{aligned} \tag{8}$$

So transversality assumption is valid with the constant $C = b^{|\Lambda|} 2^{-\sum_{i\in\Lambda}|i-i_0|}$. Note that, in order to obtain the inequality in (8), it is important that one has the parameter ε_{il} for all $(i,l) \in \Lambda \times \tilde{\Lambda}$. This is easy to understand. Let x, y be such that $x_l = y_l$ for all l except l_0. If ε_{il_0} were missing, these two points would not "see" the change of parameters. This is the reason why we have to consider the infinite dimensional parameter space E.

Remark 3.1. (1) Taking any coupled map lattice which is close to T_0 in the sense that it has an invariant set close to $\mathcal{K}$, one can repeat the above arguments without changing the conjugacy A_ε. Therefore it is possible to decompose a suitable space of coupled map lattices into leaves such that inside each leaf almost every system has infinitely many SRB-measures. This shows that the uniqueness of the SRB-measure is a very atypical situation. The explicit form of the conjugacy A_ϵ is irrelevant. It is simply enough to find one. In order to apply Theorem 2.3 it is essential that the map depends on all coordinates such that the decay rate is not faster than the one in the definition of the metric. More precisely, there has to be some lower bound on the decay so that one can define an appropriate auxiliary metric (see (8)).

(2) Note that by Theorem 2.3 the densities of $(\pi_\Lambda \circ A_\epsilon)_*\mu$ are smooth, in particular, Hölder continuous. The uniqueness proof of Bricmont and Kupiainen fails for these measures because there are regions where the density is zero, and so one cannot take the logarithm of the densities.

Finally, I would like to say a couple of words about the definition of SRB-measure. I will mainly concentrate on coupled map lattices and refer to the review of Young [61] for more general discussion. In the case of Axiom A diffeomorphisms there are several equivalent ways to define the SRB-measure.

For describing these let M be a set, $T : M \to M$ a map, and $\mathcal{L}$ some preferable measure on M (for example the Lebesgue measure if M is a manifold). First of all, SRB-measure μ is natural meaning that there exists an open set $U \subset M$ such that for all $\nu \ll \mathcal{L}|_U$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} (T_*)^i \nu = \mu$$

convergence being in the weak* topology. Secondly, SRB-measure μ is observable, that is, there exists an open set $U \subset M$ such that for $\mathcal{L}$-almost all $x \in U$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \delta_{T^i(x)} = \mu \ .$$

Thirdly, SRB-measure is an equilibrium state for an interaction constructed from the derivative of T (see Sect. 1). Fourthly, SRB-measure is (the unique) invariant measure absolutely continuous with respect to $\mathcal{L}$, or its conditional distributions on unstable leaves are absolutely continuous with respect to the corresponding Lebesgue measure. There are also other definitions for the SRB-measure like that it is a measure which is stable under random perturbations, but I will not consider these and refer to [61] for more information.

Although all the above definitions are equivalent for Axiom A diffeomorphisms they are not equivalent for a general dynamical system. For example the third and the fourth definitions are meaningful only if the system is in some sense hyperbolic. Blank and Bunimovich [6] have proved that an observable measure is always natural and an ergodic invariant natural measure which is equivalent to the Lebesgue measure is observable. Applying a result of Inoue [28] (see also [43]), they showed that there are invariant non-ergodic natural measures which are not observable. In [34] Tolonen and I constructed an ergodic invariant natural measure which is not observable. See also [51] for related results. Since these different definitions are not equivalent in general, the natural question is what is the right definition for a SRB-measure in the case of coupled map lattices. The definitions of naturalness and observability assume that there exists some preferable measure on the space. On finite dimensional manifolds the Lebesgue measure (or some smooth modification of it) has clearly a special role. So, if $X = \prod_{\mathbb{Z}^d} S^1$, one may consider the infinite product of normalized Lebesgue measures on S^1 which is a probability measure on X. However, if one modifies the Lebesgue on S^1 just a little bit and takes the product of these measures, then one obtains a measure which is singular to the product of Lebesgue measures. This indicates that there is no unique preferable measure for coupled map lattices to define naturalness or observability. For the same reason the fourth definition is not directly applicable. But keeping in mind that coupled map lattices are models for real life phenomena, one may argue that, although the real life systems are high dimensional, they are nevertheless finite dimensional. At least one can make

only finite number of observations of finite number of particles. One may also interpret a measurement as a projection so it is natural to consider the finite dimensional projections of an invariant measure. In a finite dimensional subsystem it does not matter whether one uses Lebesgue measure or some modification of it. Therefore Definition 1.1 is a very natural definition.

In statistical physics a phase transition is associated with the non-uniqueness of equilibrium states. (At the temperature 0 degrees of Celsius there are two phases – water and ice.) An interesting question is whether there could be a phase transition in coupled map lattices even for small coupling due to the infinite dimensionality of the system although for the local dynamics there were unique SRB-measure. If one takes as the definition of SRB-measure the one saying that it is an equilibrium state for a potential constructed from the derivative of the map, then the results of [10, 37] imply that there are no phase transitions. If one adopts Definition 1.1, then the results of [33] imply that there is a phase transition even for small coupling. To resolve this paradox, one should analyze the results of [33] more carefully. Indeed, these results imply that for any finite $\Lambda \subset \mathbb{Z}^d$ there is an open set $U \subset X_\Lambda$ with positive Lebesgue measure such that for $\mathcal{L}^\Lambda$-almost all $x \in U$ there are boundary conditions x_1^c and x_2^c such that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \delta_{T^i(x,x_i^c)} = \mu_i, \quad i = 1, 2$$

and $\mu_1 \neq \mu_2$. However, in order to see a measure which is different from the SRB-measure constructed by Bricmont and Kupiainen, one has to choose the boundary condition carefully. In other words, the boundary condition depends on x, and one may argue that the phase transition implied by Theorem 3.1 is not a physical one.

In statistical physics it is natural to talk about equilibrium states associated to interactions since the interaction is the primary concept. However, from the point of view of dynamical systems interaction is rather a tool than a fundamental basic quantity. Thus one would like a more dynamical definition for SRB-measure. Although being natural, Definition 1.1 has the drawback that SRB-measure is not unique for small coupling. I motivated this definition by saying that a measurement is a projection. However, projection means that one takes an average over all boundary conditions. Since one can make only finitely many measurements, one studies also finitely many boundary conditions. Thus it is perhaps better to study conditional distributions on finite dimensional subsystems and demand that they are absolutely continuous with respect to the Lebesgue measure. This approach has been taken by Keller and Zweimüller in [45]. They proved the uniqueness of SRB-measure in this sense. However, they have to assume that the coupling is unidirectional over the one dimensional lattice $\mathbb{N}$ meaning that the boundary conditions have an effect on finite subsystems but the finite subsystems have no effect on the

boundary conditions. Although this assumption is reasonable one would like to relax it.

One possible alternative is to say that μ is a SRB-measure if it is natural or observable with respect to all measures of the form $\prod_{\mathbb{Z}^d} \nu$ where ν is absolutely continuous with respect to the local Lebesgue measure. A step in this direction is taken by Bardet in [5].

As a concluding remark I emphasize that although the theory of SRB-measures for weakly coupled expanding or hyperbolic maps is already quite well understood there are still some very basic problems to be solved.

Acknowledgements

I thank Maarit Järvenpää, Mika Leikas, and Tapani Tolonen for many valuable comments.

References

1. V. Baladi: *Positive transfer operators and decay of correlations*, Advanced Series in Nonlinear Dynamics **16**, (World Scientific Publishing Co. Inc., River Edge NJ 2000) 98
2. V. Baladi: *Periodic orbits and dynamical spectra*, Ergodic Theory Dynam. Systems **18**, (1998), pp 255–292 98
3. V. Baladi, M. Degli Esposti, S. Isola, E. Järvenpää, and A. Kupiainen: *The spectrum of weakly coupled map lattices*, J. Math. Pures Appl. **77**, (1998), pp 539–584 96
4. V. Baladi and H.H. Rugh: *Floquet spectrum of weakly coupled map lattices*, Comm. Math. Phys. **220**, (2001), pp 561–582 96
5. J.-B. Bardet: *Limit theorems for coupled analytic maps*, Probab. Theory Related Fields **124**, (2002), pp 151–177 111
6. M. Blank and L. Bunimovich: *Multicomponent dynamical systems: SRB measures and phase transitions*, Nonlinearity **16**, (2003), pp 387–401 109
7. M. Blank, G. Keller, and C. Liverani: *Ruelle-Perron-Frobenius spectrum for Anosov maps*, Nonlinearity **15**, (2002), pp 1905–1973 96
8. R. Bowen: *Equilibrium states and ergodic theory of Anosov diffeomorphisms*, Springer Lecture Notes in Math. **470**, (1975). 98
9. J. Bricmont and A. Kupiainen: *Coupled analytic maps*, Nonlinearity **8**, (1995), pp 379–396 96, 100, 101
10. J. Bricmont and A. Kupiainen: *High temperature expansions and dynamical systems*, Comm. Math. Phys. **178**, (1996), pp 703–732 96, 98, 101, 110
11. L.A. Bunimovich: *Coupled map lattices: one step forward and two steps back*, Chaos, order and patterns: aspects of nonlinearity—the "gran finale" (Como, 1993), Phys. D **86**, (1995), pp 248–255 95
12. L.A. Bunimovich and Ya.G. Sinai: *Spacetime chaos in coupled map lattices*, Nonlinearity **1**, (1988), pp 491–516 95, 96
13. P. Collet: *Some ergodic properties of maps of the interval*, Dynamical systems (Temuco, 1991/1992), Travaux en Cours **52** (Hermann Paris 1996) pp 55–91 98

14. R.L. Dobrushin: *Gibbsian random fields for lattice systems with pairwise interactions*, Funct. Anal. Appl. **2**, (1968), pp 292–301 100
15. R.L. Dobrushin: *The description of a random field by means of conditional probabilities and conditions on its regularity.* Theory Prob. Appl. **13**, (1968), pp 197–224 100
16. K.J. Falconer: *Hausdorff dimension and the exceptional set of projections*, Mathematika **29**, (1982), pp 109–115 103
17. K.J. Falconer: *Geometry of Fractal Sets*, (Cambridge University Press, Cambridge 1985) 106
18. K.J. Falconer and J.D. Howroyd: *Projection theorems for box and packing dimensions*, Math. Proc. Cambridge Philos. Soc. **119**, (1996), pp 287–295 103
19. K.J. Falconer and J.D. Howroyd: *Packing dimensions of projections and dimension profiles*, Math. Proc. Cambridge Philos. Soc. **121**, (1997), pp 269–286 103
20. K.J. Falconer and P. Mattila: *The packing dimension of projections and sections of measures*, Math. Proc. Cambridge Philos. Soc. **119**, (1996), pp 695–713 103
21. K.J. Falconer and T.C. O'Neil: *Convolutions and the geometry of multifractal measures*, Math. Nachr. **204**, (1999), pp 61–82 103
22. T. Fischer and H.H. Rugh: *Transfer operators for coupled analytic maps*, Ergodic Theory Dynam. Systems **20**, (2000), pp 109–143 96
23. L. Gross: *Decay of correlations in classical lattice models at high temperature*, Comm. Math. Phys. **68**, (1979), pp 9–27 100
24. X. Hu and S.J. Taylor: *Fractal properties of products and projections of measures in* $\mathbb{R}^d$, Math. Proc. Cambridge Philos. Soc. **115**, (1994), pp 527–544 103
25. B.R. Hunt and V.Yu. Kaloshin: *How projections affect the dimension spectrum of fractal measures*, Nonlinearity **10**, (1997), pp 1031–1046 103
26. B.R. Hunt and V.Yu. Kaloshin: *Regularity of embeddings of infinite-dimensional fractal sets into finite-dimensional spaces*, Nonlinearity **12**, (1999), pp 1263–1275 103, 107
27. B.R. Hunt, T. Sauer, and J.A. Yorke: *Prevalence: a translation-invariant "almost every" for infinite dimensional spaces*, Bull. Am. Math. Soc. **27**, (1992), pp 217–238
28. T. Inoue: *Sojourn times in small neighborhoods of indifferent fixed points of one-dimensional dynamical systems*, Ergodic Theory Dynam. Systems **20**, (2000), pp 241–257 109
29. R.B. Israel: *High-temperature analyticity in classical lattice systems*, Comm. Math. Phys. **50**, (1976), pp 245–257 100
30. E. Järvenpää: *A SRB-measure for globally coupled circle maps*, Nonlinearity **10**, (1997), pp 1435–1469 96
31. E. Järvenpää: *A note on weakly coupled expanding maps on compact manifolds*, Ann. Acad. Sci. Fenn. Math. **24**, (1999), pp 511–517 96
32. E. Järvenpää and M. Järvenpää: *Linear mappings and generalized upper spectrum for dimensions*, Nonlinearity **12**, (1999), pp 475–493 103
33. E. Järvenpää and M. Järvenpää: *On the definition of SRB-measures for coupled map lattices*, Comm. Math. Phys. **220**, (2001), pp 1–12 105, 106, 107, 110
34. E. Järvenpää and T. Tolonen: *Natural ergodic measures are not always observable*, preprint 297, http://www.math.jyu.fi/tutkimus/julkaisut.html 109
35. M. Järvenpää: *On the upper Minkowski dimension, the packing dimension, and orthogonal projections*, Ann. Acad. Sci. Fenn. Math. Diss. **99**, (1994), pp 1–34 103
36. M. Jiang: *Equilibrium states for lattice models of hyperbolic type*, Nonlinearity **8**, (1995), pp 631–659 96

37. M. Jiang and Ya.B. Pesin: *Equilibrium measures for coupled map lattices: existence, uniqueness and finite-dimensional approximations*, Comm. Math. Phys. **193**, (1998), pp 675–711 96, 110
38. W. Just: *Globally coupled maps: phase transitions and synchronization*, Phys. D **81**, (1995), pp 317–340 96
39. W. Just: *Bifurcations in globally coupled map lattices*, J. Statist. Phys. **79**, (1995), pp 429–449 96
40. K. Kaneko (ed): *Theory and applications of coupled map lattices*, Nonlinear Sci. Theory Appl. (John Wiley & Sons Ltd., Chichester 1993) 95
41. R. Kaufman: *On Hausdorff dimension of projections*, Mathematika **15**, (1968), pp 153–155. 102, 103
42. R. Kaufman and P. Mattila: *Hausdorff dimension and exceptional sets of linear transformations*, Ann. Acad. Sci. Fenn. Ser. A I Math. **1**, (1975), pp 387–392 103
43. G. Keller: Completely mixing maps without limit measure, Colloq. Math. **100** (2004), no. 1, 73–76. 109
44. G. Keller and M. Künzle: *Transfer operators for coupled map lattices*, Ergodic Theory Dynam. Systems **12**, (1992), pp 297–318 96
45. G. Keller and R. Zweimüller: *Unidirectionally coupled interval maps: between dynamics and statistical mechanics*, Nonlinearity **15**, (2002), pp 1–24 110
46. M. Marstrand: *Some fundamental geometrical properties of plane sets of fractional dimension*, Proc. London Math. Soc. (3) **4**, (1954), pp 257–302 102
47. P. Mattila: *Hausdorff dimension, orthogonal projections and intersections with planes*, Ann. Acad. Sci. Fenn. Math. **1**, (1975), pp 227–244 102
48. P. Mattila: *Hausdorff dimension and capacities of intersection of sets in n-space*, Acta Math. **152**, (1984), pp 77–105 103
49. P. Mattila: *Geometry of Sets and Measures in Euclidean Spaces: Fractals and rectifiability*, (Cambridge University Press, Cambridge 1995) 97, 103, 104, 106
50. P. Mattila: *Hausdorff dimension, projections, and the Fourier transform*, Publ. Math. **48**, (2004), pp 3–48 103
51. M. Misiurewicz and A. Zdunik: Convergence of images of certain measures. In: *Statistical physics and dynamical systems* (Köszeg, 1984), pp 203–219, Progr. Phys. **10**, (Birkhäuser Boston, Boston MA 1985) 109
52. Y. Peres and W. Schlag: *Smoothness of projections, Bernoulli convolutions, and the dimension of exceptions*, Duke Math. J. **102**, (2000), pp 193–251 103, 104, 105
53. Ya.B. Pesin and Ya.G. Sinai: Space-time chaos in chains of weakly interacting hyperbolic mappings. Translated from the Russian by V.E. Nazaĭkinskiĭ. Adv. Soviet Math. **3**, In: *Dynamical systems and statistical mechanics* (Moscow, 1991), pp 165–198, (Amer. Math. Soc., Providence RI 1991) 96
54. D. Ruelle: *A measure associated with Axiom A attractors*, Amer. J. Math. **98**, (1976), pp 619–654 98
55. D. Ruelle: *Thermodynamic formalism. The mathematical structures of classical equilibrium statistical mechanics*, Encyclopedia of Mathematics and its Applications 5 (Addison-Wesley Publishing Co., Reading Mass. 1978) 100
56. T.D. Sauer and J.A. Yorke: *Are the dimensions of a set and its image equal under typical smooth functions?*, Ergodic Theory Dynam. Systems **17**, (1997), pp 941–956 103
57. M. Schmitt: *Spectral theory for nonanalytic coupled map lattices*, Nonlinearity **17**, (2004), pp 671–689 96
58. B. Simon: *The Statistical Mechanics of Lattice Gases, Vol 1*, (Princeton University Press, Princeton New Jersey 1993) 100, 101

59. Ya. Sinai: *Gibbs measures in ergodic theory*, Russian Math. Surveys **27**, (1972), pp 21–69 98
60. V. Volevich: *Construction of an analogue of the Bowen-Ruelle-Sinai measure for a multidimensional lattice of interacting hyperbolic mappings* (Russian), Mat. Sb. **184**, (1993), pp 17–36; translation in Russian Acad. Sci. Sb. Math. **79**, (1994), pp 347–363 96
61. L.-S. Young: *What are SRB measures, and which dynamical systems have them?*, J. Statist. Phys. **108**, (2002), pp 733–754 108, 109

SRB-Measures for Coupled Map Lattices

E. Järvenpää

Department of Mathematics and Statistics, P.O. Box 35 (MaD), FIN-40014
University of Jyväskylä, Finland
esaj@maths.jyu.fi

This chapter is divided into three parts. I start with a review of the existence results of SRB-measures for coupled map lattices. In the second part I give a brief introduction to the behaviour of Hausdorff dimension under general projections. Finally, I construct a counterexample to the Bricmont–Kupiainen conjecture and discuss its role in the definition of SRB-measures for coupled map lattices.

1 Existence of SRB-measures

A coupled map lattice is a discrete time dynamical system (X, T) consisting of the phase space X and the mapping $T : X \to X$. The phase space $X = \prod_L M$ is the product of a compact (finite dimensional) manifold M over some lattice L equipped with the product topology. The dynamics $T = A \circ F$ is a composition of the uncoupled map $F = \prod_L f$, where $f : M \to M$ is called the local dynamics, and the coupling $A : X \to X$, which ties the independent local systems into one infinite dimensional system. I will mainly concentrate on the situation where M is the unit circle S^1, f is an expanding smooth map, L is the d-dimensional lattice $\mathbb{Z}^d$, and the coupling A is a small perturbation of the identity such that the dependence of the ith coordinate $A(x)_i$ on x_j decreases exponentially fast with respect to the natural distance $|i-j|$ on $\mathbb{Z}^d$.

Coupled map lattices were introduced by Kaneko in 1983. I refer to [40] (see also [11]) for an overview of the numerous results obtained in the 1980's and I will concentrate on the existence results of SRB-measures (for the discussion on the definition of SRB-measure see the end of Sect. 3). The existence of SRB-measures for coupled map lattices was first proved by Bunimovich and Sinai in [12]. They studied a system where the phase space $X = \prod_{\mathbb{Z}}[0,1]$ is the product of unit intervals over the one dimensional lattice $\mathbb{Z}$, the local dynamics $f : [0,1] \to [0,1]$ is an expanding $C^{1+\delta}$ map (that is, f' is Hölder continuous with exponent δ), and A is a special nearest neighbour coupling, meaning that the value of $A(x)_i$ depends only on x_{i-1}, x_i, and x_{i+1}. They

E. Järvenpää: *SRB-Measures for Coupled Map Lattices*, Lect. Notes Phys. **671**, 95–114 (2005)
www.springerlink.com

proved that there exists an ergodic invariant Borel probability measure μ on X such that its projections μ_Λ onto finite dimensional subsystems $X_\Lambda = \prod_\Lambda [0,1]$ ($\Lambda \subset \mathbb{Z}$ is finite) are absolutely continuous with respect to the corresponding Lebesgue measure. The measure μ is also mixing with respect to the dynamics T (which I will call the time dynamics) and the space dynamics induced by the natural left shift $\sigma : \mathbb{Z} \to \mathbb{Z}$, $\sigma(i) = i - 1$. Pesin and Sinai [53] extended this result to the case where the local maps are (genuinely) hyperbolic.

The essential feature of the coupling used in [12, 53] is that it is identity on the boundaries of the space. This guarantees the existence of Markov partitions which enables the use of thermodynamic formalism, that is, one can apply the methods of statistical physics. Keller and Künzle [44] studied a more general coupling and more general local dynamics, namely only piecewise $C^{1+\delta}$, and obtained the existence of an SRB-measure. However, they managed to prove mixing properties only for finite dimensional approximations, not for the whole infinite dimensional system. For corresponding results in the case of hyperbolic local maps, see [60].

In [9] Bricmont and Kupiainen extended the aforementioned results in several ways. First of all, they studied d-dimensional lattice $\mathbb{Z}^d$ for arbitrary $d \in \mathbb{N}$. Secondly, they had an infinite range coupling which was only demanded to be small and to decay exponentially with respect to the distance between lattice sites. Thirdly, they proved good mixing properties, namely, they obtained an exponential decay of correlations for local observables in the infinite dimensional system. The price for this is that they had to consider real analytic expanding maps on the local phase space S^1. In [10] they obtained the same results for $C^{1+\delta}$ maps. The fact that similar methods can be applied if one replaces the unit circle by any compact manifold admitting expanding smooth maps was observed in [31]. Jiang [36] and Jiang and Pesin [37] have proved the corresponding results as Bricmont and Kupiainen in the case where the local dynamics is hyperbolic. Inspired by the works of Just [38, 39], I studied in [30] globally coupled maps and proved the existence of a SRB-measure. Furthermore, I verified that correlations decay like $1/N$ where N is the system size.

A natural question arising from these results is whether one can say something about the spectrum of the Perron–Frobenius operator associated to the infinite dimensional system. This question has been studied for example by Baladi et al. [3], Fischer and Rugh [22], Baladi and Rugh [4], Blank, Keller, and Liverani [7], Schmitt [57], and many others. For more information on this approach see the presentation of Gerhard Keller in this school.

I try to give a rough idea of the proofs of the results in [10] without going too deep into technical details. For readers' convenience I give explicitly the assumptions of [10]. The phase space is $X = \prod_{\mathbb{Z}^d} S^1$. It is practical to use the covering map $x_j \mapsto e^{i2\pi x_j}$ and describe the mappings via their lifts. The local dynamics is given by an orientation preserving expanding $C^{1+\delta}$ map $f : \mathbb{R} \to \mathbb{R}$. Note that this implies the existence of $\lambda > 1$ and $k \in \mathbb{N} \setminus \{0, 1\}$ such that $f'(x) > \lambda$ and $f(x + 1) = f(x) + k$ for all $x \in \mathbb{R}$. One assumes that

$f(0) \in [0,1)$. For all i the ith component $A(x)_i$ of the coupling is assumed to be continuously differentiable with respect to every variable x_j such that for all $x, y \in X$

$$\left| \frac{\partial A(x)_i}{\partial x_j} - \delta_{ij} \right| \leq \varepsilon e^{-\beta|i-j|}$$

and

$$\left| \frac{\partial A(x)_i}{\partial x_j} - \frac{\partial A(y)_i}{\partial x_j} \right| \leq \varepsilon \sum_{k \in \mathbb{Z}^d} e^{-\beta(|i-j|+|i-k|)} |x_k - y_k|^\delta$$

for some $\varepsilon, \beta, \delta > 0$. Here δ_{ij} is the Kronecker's delta. Furthermore, the coupling is assumed to be space translation invariant, that is, $A = \tau_i \circ A \circ \tau_i^{-1}$ for all $i \in \mathbb{Z}^d$. Here $(\tau_i(x))_j = x_{j-i}$.

Example 1.1. An example of a coupling satisfying the above assumptions is

$$A(x)_i = x_i + \frac{\varepsilon}{(2\pi)^2} \sum_{k \in \mathbb{Z}^d} 2^{-|i-k|} \cos(2\pi(x_i - x_k))$$

with $\beta = \log 2/2$.

Remark 1.1. In this chapter I always assume that measures are Borel regular (for the definition see [49, Definition 1.5]). Since Bricmont and Kupiainen defined a regular measure meaning something different (see Definition 1.2) I do not explicitly mention Borel regularity to avoid confusion.

Definition 1.1. *A measure μ is T-invariant if $T_*\mu = \mu$ where the image measure is defined by $T_*\mu(B) = \mu(T^{-1}(B))$ for all Borel sets B. A T-invariant probability measure μ on X is a SRB-measure if $\mu_\Lambda \ll \mathcal{L}^\Lambda$ for all finite $\Lambda \subset \mathbb{Z}^d$ where $\mathcal{L}$ is the Lebesgue measure on X. The notation $\mu \ll \nu$ means that measure μ is absolutely continuous with respect to ν. If $\mu \ll \nu$ then the Radon–Nikodym derivative $d\mu/d\nu$ is called the density of μ with respect to ν.*

For the differentiation theory of measures see [49, Sect. 2].

Definition 1.2. *Let $\Lambda \subset \mathbb{Z}^d$ be finite. The space $C^\delta(X_\Lambda)$ consists of continuous functions h on X_Λ such that*

$$|h(x) - h(y)| \leq C \sum_{i \in \Lambda} |x_i - y_i|^\delta.$$

We equip $C^\delta(X_\Lambda)$ with the norm

$$\|h\|_\delta = \|h\|_\infty + \sup_{x,y \in X_\Lambda} \frac{|h(x) - h(y)|}{\sum_{i \in \Lambda} |x_i - y_i|^\delta} \ .$$

A probability measure μ on X is regular if $\mu_\Lambda \ll \mathcal{L}^\Lambda$, its density h_Λ is continuous and strictly positive, and $\log h_\Lambda \in C^\delta(X_\Lambda)$ for all finite $\Lambda \subset \mathbb{Z}^d$.

In [10] the following theorem was proved.

Theorem 1.1. *Let T satisfy the assumptions given above. There exists $\varepsilon_0 > 0$ such that for all $\varepsilon < \varepsilon_0$ the map T has a SRB-measure μ. Furthermore, μ is invariant and exponentially mixing under the space-time translations generated by τ_i and T: there exists $m > 0$ and $C < \infty$ such that for all finite $B, D \subset \mathbb{Z}^d$ and for all $g \in L^\infty(X_B)$ and $h \in C^\delta(X_D)$ we have*

$$\left| \int g \circ T^n h d\mu - \int g d\mu \int h d\mu \right| \leq C \ \|g\|_\infty \ \|h\|_\delta \ e^{-m(n+d(B,D))}$$

where $d(B, D) = \min\{|i - j| \mid i \in B, j \in D\}$ and C depends on the diameters of B and D. Finally, for all regular measures ν the sequence $(T_*)^n \nu$ tends to μ in the weak* topology.

There are three main tools in the proof of Theorem 1.1: Perron–Frobenius operator, equilibrium states associated to interactions in lattice models of statistical physics, and polymer expansions.

Definition 1.3. *Let M be a compact manifold, $\mathcal{L}$ the Lebesgue measure on M, and $T : M \to M$ a non-singular mapping meaning that $\mathcal{L}(B) = 0$ if and only if $\mathcal{L}(T^{-1}(B)) = 0$ for Borel sets B. The Perron–Frobenius operator $P : L^1(M) \to L^1(M)$ associated to T is defined by the equation*

$$\int g \circ T \ h \, d\mathcal{L} = \int g P(h) \, d\mathcal{L} \tag{1}$$

for all $g \in L^\infty(M)$ and $h \in L^1(M)$.

Observe that if $\mu \ll \mathcal{L}$ then also $T_*\mu \ll \mathcal{L}$ by the non-singularity of T. Thus if h is the Radon–Nikodym derivative of μ with respect to $\mathcal{L}$ then $P(h)$ is the Radon–Nikodym derivative of $T_*\mu$ with respect to $\mathcal{L}$. (In the sequel the Radon–Nikodym derivative is called simply the density.) Therefore, P may be viewed as the restriction of T_* to measures which are absolutely continuous with respect to $\mathcal{L}$. From this it is obvious that if h is a fixed point of P, that is, it is an eigenstate with eigenvalue 1, then $h\mathcal{L}$ is an invariant measure for T. It is a well-known result that if T is an expanding $C^{1+\delta}$ map on a compact manifold M then the spectrum of P, restricted to the space of functions of bounded variation, consists of a simple eigenvalue 1 and the rest of the spectrum is inside a disc of radius γ for some $\gamma < 1$. This implies that any smooth initial density will converge under the iteration of P to the density of the unique invariant measure μ which is absolutely continuous with respect to $\mathcal{L}$. In fact, the SRB-measure μ is equivalent with $\mathcal{L}$. These results are due to Sinai [59], Ruelle [54], and Bowen [8]. For more information on these questions see [1, 2]. An easy-to-access exposition of interval maps is [13].

I will need the following explicit form of the Perron–Frobenius operator which is valid in the expanding orientation preserving case:

$$P(h)(x) = \sum_{y:T(y)=x} \frac{h(y)}{\det DT(y)} \cdot \tag{2}$$

This formula is easily verified via the change of variables.

The idea of the proof of Theorem 1.1 is that one first takes a finite $\Lambda \subset \mathbb{Z}^d$, fixes some boundary condition x_{Λ^c} on X_{Λ^c} (where $\Lambda^c = \mathbb{Z}^d \setminus \Lambda$), and obtains in this way an expanding map on the finite dimensional manifold X_Λ. More precisely, $T_\Lambda : X_\Lambda \to X_\Lambda$ is defined by $T_\Lambda = A_\Lambda \circ \prod_\Lambda f$, where $A_\Lambda(x)_i = A(x_\Lambda, x_{\Lambda^c})_i$ for all $i \in \Lambda$ and the boundary condition x_{Λ^c} does not evolve in time. Then one iterates the constant function 1 with the Perron–Frobenius operator P_Λ associated to this finite dimensional approximation. By the aforementioned results concerning finite dimensional expanding maps the sequence $P_\Lambda^N(1)$ will tend to the density of the unique SRB-measure μ_Λ of the finite dimensional system. Finally, letting Λ tend to $\mathbb{Z}^d$ along some increasing sequence, one shows that the sequence μ_Λ converges to some measure μ which is independent of the boundary condition x_{Λ^c}. This limiting measure is the SRB-measure of Theorem 1.1.

As a result of the above method one obtains the following lattice gas model. In addition to the lattice $\mathbb{Z}^d$ each iterate of the Perron–Frobenius operator will give a new copy of $\mathbb{Z}^d$ leading to the lattice $L = \mathbb{Z}^d \times \mathbb{N}$. As a result we have a phase space $\Omega = \prod_L M_{(i,j)}$, where $M_{(i,0)} = S^1$ for all $i \in \mathbb{Z}^d$ and $M_{(i,j)} = \{0, \dots, k-1\}$ for all $i \in \mathbb{Z}^d$ and $j \in \mathbb{N} \setminus \{0\}$. Here k is the number of preimages of any point for the map $f : S^1 \to S^1$. Denoting by f also the lift of f, one sees that the different preimages of a point $x \in [0,1)$ are of the form $f^{-1}(x+s)$ where $s = 0, 1, \dots, k-1$. Let $\psi_s(x)_i = f^{-1}(A_\Lambda^{-1}(x)_i + s_i)$ where $s \in \{0, \dots, k-1\}^\Lambda$. Then by (2)

$$P_\Lambda(h)(x) = \det DA_\Lambda^{-1}(x) \sum_{s \in \{0,\dots,k-1\}^\Lambda} \frac{h(\psi_s(x))}{\prod_{i\in\Lambda} f'(\psi_s(x)_i)}$$

and further

$$P_\Lambda^N(1)(x) = \sum_{s^1,\dots,s^N} \prod_{t=1}^{N} \Big(\det DA_\Lambda^{-1}(\psi_{s^{t-1}} \circ \cdots \circ \psi_{s^1}(x)) \prod_{i\in\Lambda} (f'(\psi_{s^t} \circ \cdots \circ \psi_{s^1}(x)_i))^{-1} \Big) \cdot$$

One begins the proof by writing

$$P_\Lambda^N(1) = e^{-H} \tag{3}$$

where H is a real valued function on $\Omega_{\Lambda\times\{0,\dots,N\}}$. In the terminology of statistical physics the function H is called Hamiltonian. In order to apply the methods of statistical physics one still has to localize the Hamiltonian, meaning that one needs to find an interaction Φ such that

$$H = \sum_{Y \subset \Lambda \times \{0,\dots,N\}} \Phi_Y$$

where Φ_Y is a function on Ω_Y. One proves that the interaction Φ_Y, which depends on Λ, N, and the chosen boundary condition x_{Λ^c}, will tend to an interaction Φ_Y, which is independent of Λ, N, and the boundary condition x_{Λ^c}, as N tends to infinity and Λ tends to $\mathbb{Z}^d$. The results of statistical mechanics guarantee that, provided the limiting interaction Φ has some nice properties, there will be a unique equilibrium state $\tilde{\mu}$ associated to it. This equilibrium state is a measure on Ω and it will be exponentially mixing with respect to translations in L. Note that the translation in $\mathbb{N}$ direction corresponds to the time dynamics of the original coupled map lattice. Thus the projection of $\tilde{\mu}$ onto $\Omega_{\mathbb{Z}^d \times \{0\}}$ will be the SRB-measure of Theorem 1.1.

An excellent introduction and a handbook of lattice gas models of statistical physics is [58]. A more compressed but still complete exposition can be found in [55]. I refer to these books for more information on equilibrium and Gibbs states and I briefly explain some relevant results concerning interactions. In the case where the underlying lattice is at least two dimensional, it is important to distinguish between several norms used in the space of interactions. For simplicity I assume that the interactions are translation invariant and 0 is a fixed point in the lattice. If

$$\|\Phi\|_0 = \sum_{Y \ni 0} |\Phi_Y|$$

is finite, then there exists an equilibrium state (see [58, Corollary III.2.9]). Assuming that

$$\|\Phi\|_1 = \sum_{Y \ni 0} |Y||\Phi_Y|$$

is small, the equilibrium state is unique (see [14, 15]). If for some $\gamma > 0$

$$\|\Phi\|_2 = \sum_{Y \ni 0} e^{\gamma d(Y)} |\Phi_Y|$$

is small, then the equilibrium state is unique and the correlations decay exponentially (see [23]). Finally, supposing that for some $\gamma > 0$

$$\|\Phi\|_3 = \sum_{Y \ni 0} e^{\gamma |Y|} |\Phi_Y|$$

is small, the equilibrium state is unique and the correlation functions are analytic (see [29]). Here $|Y|$ is the number of elements in Y and $d(Y)$ is the diameter of Y.

The interaction rising from the proof of Theorem 1.1 does not fall into any of the above classes. In fact, it is of the form $\Phi^0 + \Phi^1$ where Φ^0 is a finite range one dimensional interaction and $\|\Phi^1\|_2$ is small. In [9] the analyticity

assumptions guarantee that $\|\Phi^1\|_3$ is small. Since the behaviour of one dimensional lattice models is radically different from higher dimensional ones, this kind of interaction will imply the nice properties stated in Theorem 1.1.

How does one localize a Hamiltonian? One method (also used by Bricmont and Kupiainen) is telescoping. Namely, let $H = H(x_1, \ldots, x_n)$ and 0 be some fixed value which every coordinate may attain. Then

$$\begin{aligned} H(x_1, \ldots, x_n) =& H(x_1, \ldots, x_n) - H(x_1, \ldots, x_{n-1}, 0) \\ &+ H(x_1, \ldots, x_{n-1}, 0) - H(x_1, \ldots, x_{n-2}, 0, 0) \\ &+ \cdots + H(x_1, 0, \ldots, 0) - H(0, \ldots, 0) + H(0, \ldots, 0) \ . \end{aligned}$$

Now one can define

$$\Phi_{\{1,\ldots,k\}}(x_1, \ldots, x_k) = H(x_1, \ldots, x_k, 0, \ldots, 0) - H(x_1, \ldots, x_{k-1}, 0, \ldots, 0)$$

and set $\Phi_Y \equiv 0$ if $Y \neq \{1, \ldots, k\}$ for all $k = 1, \ldots, n$. To use this method, some numbering of points in $\Lambda \times \{0, \ldots, N\}$ has to be fixed. The ε in the coupling guarantees that if the set Y contains points with different space coordinates, then there will be ε in the interaction. The exponential decay of the coupling, in turn, is needed to make the norm $\|\cdot\|_2$ small. Finally, one has to decompose the Hamiltonian also in the time direction. Note that in time direction one has a sequence of ψ's (see (3)). Each ψ is a contraction with coefficient $\lambda' < 1$ since f^{-1} is a contraction and the coupling is small. Although λ' is not necessarily small, $(\lambda')^M$ is small for M large. So one has to divide the terms into two groups depending on the number of ψ's they contain. This is the reason why one has the Φ^0-part in the interaction which is not small.

Since the interaction does not fall into any of the well-known classes of interactions described above, the claims of Theorem 1.1 have to be proved explicitly. It is clear that if there is some correlation between observables supported on B and D there must be interactions connecting the sets B and D. There are two possibilities: either there is one term Φ_Y connecting these sets, or there is a chain of terms $\Phi_{Y_1}, \ldots, \Phi_{Y_n}$ starting from B and ending in D. In the first case, the smallness of correlation is due to the exponential decay of the interaction with respect to the diameter of the set. In the second case, there must be many terms and each of them contains one ε. To make this argument rigorous one has to have a good way of indexing all different possibilities. The polymer expansion is a technical tool which is developed for these purposes. It enables one to write an explicit formula for the correlation functions and expectations of local observables, that is, functions depending only on finitely many coordinates. From this formula one finds an upper bound for the rate of mixing. I refer to [9, 10] for more details. An excellent introduction to polymer expansions is [58, Sect. V.7].

In [9] Bricmont and Kupiainen made a conjecture that there is only one SRB-measure for the systems they considered. Note that according to Theorem 1.1 this conjecture is true in the class of regular measures (see Definition 1.2).

2 Projection Results

In this section I give an overview of some results in the field of geometric measure theory which seem to be totally unrelated to the results of the previous section. However, it appears that they are closely connected to the solution of Bricmont–Kupiainen conjecture.

I start with a simple example.

Example 2.1. Consider a line segment in $\mathbb{R}^3$. If one projects it onto a two dimensional plane, one obtains a line segment, unless the plane happens to be perpendicular to the original line segment in which case one obtains a point. Hence this one dimensional set is projected generically onto a one dimensional set. If, instead of line segment, one projects a cube the result is two dimensional. Thus the projection of the three dimensional cube is two dimensional.

Amazingly these two simple examples illustrate the general behaviour of Hausdorff dimensions of projections as the following theorem states:

Theorem 2.1. *Let $0 < m < n$ be integers and $B \subset \mathbb{R}^n$. Then*

$$\dim_{\mathrm{H}}(P_V(B)) = \min\{\dim_{\mathrm{H}}(B), m\}$$

for $\gamma_{n,m}$-almost all $V \in G(n,m)$. Here $G(n,m)$ is the Grasmann manifold of m-planes in $\mathbb{R}^n$, $\gamma_{n,m}$ is the Haar measure on $G(n,m)$, P_V is the orthogonal projection onto V, and $\dim_{\mathrm{H}}(B)$ is the Hausdorff dimension of B.

This theorem was first proven by Marstrand [46] in the plane, then by Kaufman [41] in the plane using different methods, and finally by Mattila [47] in full generality. The corresponding result is true also for measures. Recall that the definition of the Hausdorff dimension of a measure is as follows:

Definition 2.1. *The Hausdorff dimension of a finite measure μ on $\mathbb{R}^n$ is*

$$\begin{aligned}\dim_{\mathrm{H}}(\mu) &= \inf\{\dim_{\mathrm{H}}(B) \mid B \text{ is a Borel set with } \mu(B) > 0\} \\ &= \mu\text{-}\operatorname*{ess\,inf}_{x\in\mathbb{R}^n} \underline{\dim}_{\mathrm{loc}}\,\mu(x)\end{aligned}$$

where

$$\underline{\dim}_{\mathrm{loc}}\,\mu(x) = \liminf_{r\to 0} \frac{\log \mu(B(x,r))}{\log r}$$

is the lower local dimension of μ at x. Here $B(x,r)$ is the closed ball centred at x with radius r.

The analogue of Theorem 2.1 is valid for measures:

Theorem 2.2. *Let $0 < m < n$ be integers and let μ be a finite measure on $\mathbb{R}^n$. Then for $\gamma_{n,m}$-almost all $V \in G(n,m)$*

$$\begin{aligned}\dim_{\mathrm{H}}(P_{V*}\mu) = \dim_{\mathrm{H}}(\mu), &\text{ if } \dim_{\mathrm{H}}(\mu) \le m \\ P_{V*}\mu \ll \mathcal{L}^m, &\text{ if } \dim_{\mathrm{H}}(\mu) > m\,.\end{aligned}$$

Theorem 2.2 is more or less explicitly stated in [20, 24, 41, 48]. Theorems 2.1 and 2.2 may be summed up by saying that Hausdorff dimension is preserved under typical projections. Note that the set of exceptional directions may be quite large although it has zero measure (see [16, 42, 52]).

There are several extensions of these theorems. Instead of Hausdorff dimension, one may consider different dimensions like packing dimension, box counting dimension, multifractal q-dimensions etc. (see [18, 19, 20, 21, 25, 32, 35]). Note that the choice of dimension effects also on the results obtained. One may also replace typical projections by a prevalent set of C^1-maps (see [25, 32, 56]). Hunt and Kaloshin [26] showed that the corresponding results are not true in infinite dimensional situations. I refer to the excellent survey of Mattila [50] for more information.

To give some flavour of the aforementioned results, I make a simple calculation which contains some essential features related to projections. One crucial ingredient in this context is the s-energy of a measure which I now define.

Definition 2.2. *Let μ be a finite measure on a metric space (X, d) and $s > 0$. The s-energy of μ is*

$$I_s(\mu) = \iint d(x,y)^{-s} d\mu(x) d\mu(y) \, .$$

The finiteness of s-energy is closely related to the Hausdorff dimension of a measure. In fact, if $I_s(\mu) < \infty$ then $\dim_{\mathrm{H}}(\mu) \geq s$. On the other hand, assuming that $s < \dim_{\mathrm{H}}(\mu)$, the measure μ has a restriction with finite s-energy (see [49, pp. 109–110]). This indicates the relevance of the energies of projected measures when calculating their Hausdorff dimensions. To begin with I derive a useful formula for integrals according to which for any measure μ on a separable metric space X

$$\int f(x) d\mu(x) = \int_0^\infty \mu(\{x \in X \mid f(x) \geq t\}) \; d\mathcal{L}^1(t) \tag{4}$$

for all non-negative Borel functions f (see [49, Theorem 1.15]). Indeed, letting $A = \{(x,t) \in X \times \mathbb{R} \mid f(x) \geq t\}$ we have

$$\begin{aligned}
&\int_0^\infty \mu(\{x \in X \mid f(x) \geq t\}) \; d\mathcal{L}^1(t) = \int_0^\infty \mu(\{x \in X \mid (x,t) \in A\}) \; d\mathcal{L}^1(t) \\
&= \int \mathcal{L}^1(\{t \in [0,\infty) \mid (x,t) \in A\}) \; d\mu(x) = \int \mathcal{L}^1([0, f(x)]) \; d\mu(x) \\
&= \int f(x) \; d\mu(x).
\end{aligned}$$

Let μ be a measure on $\mathbb{R}^n$. The starting point of the energy calculations of projections of μ is [49, Lemma 3.11] according to which there exists a constant c such that

$$\gamma_{n,m}(\{V \in G(n,m) \mid |P_V(x)| \le \delta\}) \le c\delta^m |x|^{-m}. \tag{5}$$

Combining this with (4), we obtain for all $0 < s < m$ and $x \in \mathbb{R}^n \setminus \{0\}$ the following result (see [49, Corollary 3.12]):

$$\begin{aligned}
&\int |P_V(x)|^{-s} d\gamma_{n,m}(V) = \int_0^\infty \gamma_{n,m}(\{V \in G(n,m) \mid |P_V(x)|^{-s} \ge t\})\, d\mathcal{L}^1(t) \\
&= \int_0^\infty \gamma_{n,m}(\{V \in G(n,m) \mid |P_V(x)| \le t^{-\frac{1}{s}}\})\, d\mathcal{L}^1(t) \\
&\le \int_0^{|x|^{-s}} d\mathcal{L}^1(t) + \int_{|x|^{-s}}^\infty \gamma_{n,m}(\{V \in G(n,m) \mid |P_V(x)| \le t^{-\frac{1}{s}}\})\, d\mathcal{L}^1(t) \\
&\le |x|^{-s} + c|x|^{-m} \int_{|x|^{-s}}^\infty t^{-\frac{m}{s}}\, d\mathcal{L}^1(t) \\
&= \left(1 + \frac{cs}{m-s}\right) |x|^{-s}\,.
\end{aligned} \tag{6}$$

Using (6), we find a constant C such that

$$\begin{aligned}
\int I_s(P_{V*}\mu) d\gamma_{n,m}(V) &= \iiint |x-y|^{-s} dP_{V*}\mu(x) dP_{V*}\mu(y) d\gamma_{n,m}(V) \\
&= \iiint |P_V(x-y)|^{-s} d\mu(x) d\mu(y) d\gamma_{n,m}(V) \\
&= \iiint |P_V(x-y)|^{-s} d\gamma_{n,m}(V) d\mu(x) d\mu(y) \le C I_s(\mu).
\end{aligned} \tag{7}$$

According to (7), if $I_s(\mu) < \infty$ then $I_s(P_{V*}\mu) < \infty$ for $\gamma_{n,m}$-almost all V. This is one of the key observations of the proofs of Theorems 2.1 and 2.2. It is clear that this method will give results only for almost all projections. However, as I mentioned earlier, "almost all"-results are the best possible ones in this context.

A very important extension of the projection results is due to Peres and Schlag [52]. They replace projections by a general parametrized transversal family of mappings from a compact metric space to $\mathbb{R}^m$. Intuitively, transversality means that when the parameter is also changed the mapping is changed fast enough. Since I will use [52, Theorem 7.3] in Sect. 3, I state it explicitly but not in the full generality.

Definition 2.3. *Let (X,d) be a compact metric space, $Q \subset \mathbb{R}^n$ an open connected set, and $\Pi : Q \times X \to \mathbb{R}^m$ a continuous map with $n \ge m$. Define for all $x \ne y \in X$*

$$\Phi_{x,y}(\lambda) = \frac{\Pi(\lambda,x) - \Pi(\lambda,y)}{d(x,y)}\,.$$

The mapping Π is regular if for any multi-index $\eta = (\eta_1, \dots, \eta_n) \in \mathbb{N}^n$ there exists a constant C_η such that

$$|\partial^\eta \Pi(\lambda, x)| \le C_\eta \text{ and } |\partial^\eta \Phi_{x,y}(\lambda)| \le C_\eta$$

for all $\lambda \in Q$ *and for all* $x \neq y \in X$. *Here* $\partial^\eta = \frac{\partial^{|\eta|}}{(\partial\lambda_1)^{\eta_1}\ldots(\partial\lambda_n)^{\eta_n}}$ *and* $|\eta| = \sum_{i=1}^n \eta_i$.

Definition 2.4. *A regular mapping* $\Pi : Q \times X \to \mathbb{R}^m$ *is transversal if there exists a constant* C *such that for all* $\lambda \in Q$ *and for all* $x \neq y \in X$ *the condition* $|\Phi_{x,y}(\lambda)| \le C$ *implies*

$$\det(D\Phi_{x,y}(\lambda)(D\Phi_{x,y}(\lambda))^T) \ge C^2.$$

Here the derivative with respect to λ *is denoted by* D *and* A^T *is the transpose of a matrix* A.

The following theorem from [52] gives a relation between Sobolev norms of images of measures under transversal family of regular mappings and energies of original measures.

Theorem 2.3. *Let* $\Pi : Q \times X \to \mathbb{R}^m$ *be transversal and let* μ *be a finite measure on* X *such that* $I_s(\mu) < \infty$ *for some* $s > 0$. *Then there exists a constant* C_γ *such that*

$$\int_Q \|\Pi_*\mu\|_{2,\gamma}^2 \, d\mathcal{L}^n(\lambda) \le C_\gamma I_s(\mu)$$

provided that $m + 2\gamma \le s$. *Here* $\|\cdot\|_{2,\gamma}$ *is the Sobolev norm, that is,*

$$\|\nu\|_{2,\gamma}^2 = \int_{\mathbb{R}^m} |\hat{\nu}(\xi)|^2 |\xi|^{2\gamma} d\mathcal{L}^m(\xi)$$

for any finite compactly supported measure on $\mathbb{R}^m$, *where*

$$\hat{\nu}(\xi) = \int_{\mathbb{R}^m} e^{-i\xi\cdot x} d\nu(x)$$

is the Fourier transform of ν.

Proof. See [52, Theorem 7.3].

Remark 2.1. Let ν be a finite compactly supported measure on $\mathbb{R}^n$. If $\|\nu\|_{2,0} < \infty$ then ν is absolutely continuous with respect to the Lebesgue measure $\mathcal{L}^n$ and its Radon-Nikodym derivative is L^2-integrable (see [33, Remark 2.6]).

I stated only the part of [52, Theorem 7.3] concerning measures whose dimensions are larger than m since I need only that part in Sect. 3. There is also the part which says that the dimension is preserved if it is less than m. I say briefly something about the assumptions of Theorem 2.3. It is quite obvious that some kind of smoothness is needed and almost as obvious that the regularity assumption I formulated in Definition 2.3 is not the optimal

one. The role of transversality may be explained as follows. Recall from the results of orthogonal projections that there may always be some exceptional directions. Since Theorem 2.3 deals with a parametrized family of maps one could smoothly parametrize one exceptional projection with as many parameters as one likes. Thus there must be some condition guaranteeing that if the parameter is changed also the mapping is changed. In the proof there will be a calculation like (7) in one form or the other where one needs an analogue of (5). Transversality, in turn, is needed for this.

3 Counterexample to Bricmont–Kupiainen Conjecture

In this section I describe how one can construct a counterexample to the Bricmont–Kupiainen conjecture (see [33]). This counterexample is based on the results of Sect. 2.

Consider the standard $\frac{1}{3}$-Cantor set K on the unit interval and let $\mathcal{H}^s|_K$ be the restriction of the s-dimensional Hausdorff measure to K where $s = \log 2/\log 3$ is the Hausdorff dimension of K. Then $\mathcal{H}^s|_K$ is an ergodic invariant measure for the map $f : [0,1] \to [0,1]$, $f(x) = 3x \mod 1$. Since the packing and Hausdorff dimensions of K coincide, the Hausdorff dimension of the n-fold product K^n is equal to $n \log 2/\log 3$ (see [49, Theorem 8.10]). Thus for any m one can find N such that for all $n > N$ the dimension of K^n is larger than m. This is true also for the product measure $(\mathcal{H}^s|_K)^n$. According to Theorem 2.2, the projection of $(\mathcal{H}^s|_K)^n$ is absolutely continuous with respect to $\mathcal{L}^m$ for typical m-planes. However, the Bricmont–Kupiainen conjecture deals with specific projections, namely, those determined by coordinate planes. And clearly in this example the coordinate planes are not typical. The idea of our example is that a small coupling will perturb the uncoupled map slightly such that the coordinate planes become typical ones. Actually we do not prove quite this but instead we find a parametrized family of conjugating maps such that almost all conjugates of the uncoupled map have infinitely many SRB-measures.

Let $X = \prod_{\mathbb{Z}^d} S^1$, $\mathcal{K} = \prod_{\mathbb{Z}^d} K$, and $\mu = \prod_{\mathbb{Z}^d} \mathcal{H}^s|_K$. Since $\mathcal{H}^s(K) = 1$ (see [17, Theorem 1.14]), μ is a probability measure. I denote by μ_Λ the projection of μ onto X_Λ. As in Sect. 1, I will use the same symbol for maps (like $f : S^1 \to S^1$, $f(z) = z^3$) and their lifts (like $f : \mathbb{R} \to \mathbb{R}$, $f(x) = 3x$). Consider $A_\epsilon : X \to X$,

$$A_\varepsilon(x)_i = x_i + \sum_{l \in \mathbb{Z}^d} \varepsilon_{il} 2^{-|i-l|} g(x_l)$$

where for some $\varepsilon_0 > 0$ one has $\varepsilon_{ij} \in (-\varepsilon_0, \varepsilon_0)$ for all $i, j \in \mathbb{Z}^d$, and $g : \mathbb{R} \to \mathbb{R}$ is smooth and 1-periodic. For convenience assume that $|g(x)| < 1$ for all $x \in \mathbb{R}$. It is not difficult to see that A_ε is invertible provided ε_0 is small enough (depending on $|g'|$). Namely, local invertibility follows from the

implicit function theorem, and global invertibility from the fact that A_ε is a lift of a map on X. Set $E = \prod_{\mathbb{Z}^d \times \mathbb{Z}^d}(-\varepsilon_0, \varepsilon_0)$ and $\mathcal{L} = \prod_{\mathbb{Z}^d \times \mathbb{Z}^d} \rho$ where ρ is the normalized restriction of the Lebesgue measure to $(-\varepsilon_0, \varepsilon_0)$. I use the abbreviations $E_{\Lambda\times\tilde\Lambda}$ and $\mathcal{L}^{\Lambda\times\tilde\Lambda}$ for the natural restrictions. For $\Lambda \subset \tilde\Lambda$, I denote the natural projection from $X_{\tilde\Lambda}$ onto X_Λ by $\pi_{\tilde\Lambda,\Lambda}$. Let $T_\varepsilon = A_\varepsilon \circ \prod_{\mathbb{Z}^d} f \circ A_\varepsilon^{-1}$. Now $A_{\varepsilon *}\mu$ is clearly T_ε-invariant. In [33] we proved the following theorem.

Theorem 3.1. *For $\mathcal{L}$-almost all $\varepsilon \in E$ the map T_ε has infinitely many SRB-measures.*

The proof is divided into two steps. First we fix finite $\Lambda \subset \tilde\Lambda \subset \mathbb{Z}^d$ such that $|\tilde\Lambda|s > |\Lambda|$ and consider the restriction of A to $\tilde\Lambda$ with open boundary condition, that is,

$$A_{\varepsilon,\tilde\Lambda}(x)_i = x_i + \sum_{l\in\tilde\Lambda} \varepsilon_{il} 2^{-|i-l|} g(x_l) \; .$$

Using Theorem 2.3, we show that for $\mathcal{L}^{\Lambda\times\tilde\Lambda}$-almost all $\varepsilon \in E_{\Lambda\times\tilde\Lambda}$ the projection $(\pi_{\tilde\Lambda,\Lambda} \circ A_{\varepsilon,\tilde\Lambda})_* \mu_{\tilde\Lambda}$ is absolutely continuous. For this we have to restrict the map g by demanding that $|g'| > b > 0$ in the set K. (Note that by 1-periodicity there must be points where $g' = 0$.) This guarantees that the transversality condition is satisfied. The final step is to let $\tilde\Lambda$ tend to $\mathbb{Z}^d$. This is a technical calculation and I refer to [33] for more details. Intuitively this limiting process should not cause any problems. Indeed, according to Theorem 2.3, the larger the dimension of the measure, the smoother its projection is. Thus tilting the measure also in the complement of $\tilde\Lambda$ should make the situation better. However, one should keep in mind the result of Hunt and Kaloshin [26] according to which the Hausdorff dimension may decrease under projections from infinite dimensional space onto finite dimensional subspaces. Although it is possible that everything fails at the infinite limit, our setup indicates that this is not very likely and the smoothness argument given by Theorem 2.3 is stronger. The technical calculation in [33, Proposition 3.3] shows that this is indeed the case.

Finally, the above consideration implies that there are at least two SRB-measures, namely $A_{\varepsilon *}\mu$ and the SRB-measure constructed by Bricmont and Kupiainen. However, instead of taking the standard Cantor set where one deletes the middle third, one may delete the first or the last third. At every coordinate direction one may choose one of these three sets. Clearly it is possible to choose g such that $|g'| > b > 0$ on all these three sets. In this way one obtains infinitely many SRB-measures.

Now I explain some details of the proof to illustrate what is going on. To use Theorem 2.3, let $\Pi : E_{\Lambda\times\tilde\Lambda} \times Y_{\tilde\Lambda} \to X_\Lambda$, $\Pi(\varepsilon, x) = \pi_{\tilde\Lambda,\Lambda} \circ A_{\varepsilon,\tilde\Lambda}(x)$. Here $Y_{\tilde\Lambda} = \prod_{\tilde\Lambda}[-t_0, t_0]$ for some $\frac{1}{9} < t_0 < \frac{1}{6}$. This is just a technical detail to guarantee that $|g(t) - g(t')| \geq b|t - t'|$. Now Π is clearly regular so it is enough to study the validity of the transversality assumption. We equip $Y_{\tilde\Lambda}$ with the metric

$$d(x,y)^2 = \sum_{l\in\tilde{\Lambda}} 2^{-2|i_0-l|}|x_l - y_l|^2$$

where $i_0 \in \Lambda$ is some fixed point. Now

$$\Phi_{x,y}(\varepsilon)_i = \frac{\Pi(\varepsilon,x) - \Pi(\varepsilon,y)}{d(x,y)} = \frac{x_i - y_i + \sum_{l\in\tilde{\Lambda}} \varepsilon_{il} 2^{-|i-l|}(g(x_l) - g(y_l))}{d(x,y)}$$

for all $i \in \Lambda$. Fix $i \in \Lambda$, $k = (k_1, k_2) \in \Lambda \times \tilde{\Lambda}$, and $x \neq y \in Y_{\tilde{\Lambda}}$. Then

$$D\Phi_{x,y}(\varepsilon)_{i,k} = \frac{\delta_{i,k_1} 2^{-|i-k_2|}(g(x_{k_2}) - g(y_{k_2}))}{d(x,y)} \ .$$

Thus for $i, j \in \Lambda$

$$\begin{aligned}(D\Phi_{x,y}(\epsilon)D\Phi_{x,y}(\epsilon)^T)_{i,j} =& \frac{\delta_{i,j}}{d(x,y)^2} \sum_{l\in\tilde{\Lambda}} 2^{-|i-l|-|j-l|}(g(x_l) - g(y_l))^2 \\ &\geq \delta_{i,j} b^2 2^{-|i-i_0|-|j-i_0|}.\end{aligned} \tag{8}$$

So transversality assumption is valid with the constant $C = b^{|\Lambda|} 2^{-\sum_{i\in\Lambda}|i-i_0|}$. Note that, in order to obtain the inequality in (8), it is important that one has the parameter ε_{il} for all $(i,l) \in \Lambda \times \tilde{\Lambda}$. This is easy to understand. Let x, y be such that $x_l = y_l$ for all l except l_0. If ε_{il_0} were missing, these two points would not "see" the change of parameters. This is the reason why we have to consider the infinite dimensional parameter space E.

Remark 3.1. (1) Taking any coupled map lattice which is close to T_0 in the sense that it has an invariant set close to $\mathcal{K}$, one can repeat the above arguments without changing the conjugacy A_ε. Therefore it is possible to decompose a suitable space of coupled map lattices into leaves such that inside each leaf almost every system has infinitely many SRB-measures. This shows that the uniqueness of the SRB-measure is a very atypical situation. The explicit form of the conjugacy A_ϵ is irrelevant. It is simply enough to find one. In order to apply Theorem 2.3 it is essential that the map depends on all coordinates such that the decay rate is not faster than the one in the definition of the metric. More precisely, there has to be some lower bound on the decay so that one can define an appropriate auxiliary metric (see (8)).

(2) Note that by Theorem 2.3 the densities of $(\pi_\Lambda \circ A_\epsilon)_*\mu$ are smooth, in particular, Hölder continuous. The uniqueness proof of Bricmont and Kupiainen fails for these measures because there are regions where the density is zero, and so one cannot take the logarithm of the densities.

Finally, I would like to say a couple of words about the definition of SRB-measure. I will mainly concentrate on coupled map lattices and refer to the review of Young [61] for more general discussion. In the case of Axiom A diffeomorphisms there are several equivalent ways to define the SRB-measure.

For describing these let M be a set, $T : M \to M$ a map, and $\mathcal{L}$ some preferable measure on M (for example the Lebesgue measure if M is a manifold). First of all, SRB-measure μ is natural meaning that there exists an open set $U \subset M$ such that for all $\nu \ll \mathcal{L}|_U$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} (T_*)^i \nu = \mu$$

convergence being in the weak* topology. Secondly, SRB-measure μ is observable, that is, there exists an open set $U \subset M$ such that for $\mathcal{L}$-almost all $x \in U$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \delta_{T^i(x)} = \mu \ .$$

Thirdly, SRB-measure is an equilibrium state for an interaction constructed from the derivative of T (see Sect. 1). Fourthly, SRB-measure is (the unique) invariant measure absolutely continuous with respect to $\mathcal{L}$, or its conditional distributions on unstable leaves are absolutely continuous with respect to the corresponding Lebesgue measure. There are also other definitions for the SRB-measure like that it is a measure which is stable under random perturbations, but I will not consider these and refer to [61] for more information.

Although all the above definitions are equivalent for Axiom A diffeomorphisms they are not equivalent for a general dynamical system. For example the third and the fourth definitions are meaningful only if the system is in some sense hyperbolic. Blank and Bunimovich [6] have proved that an observable measure is always natural and an ergodic invariant natural measure which is equivalent to the Lebesgue measure is observable. Applying a result of Inoue [28] (see also [43]), they showed that there are invariant non-ergodic natural measures which are not observable. In [34] Tolonen and I constructed an ergodic invariant natural measure which is not observable. See also [51] for related results. Since these different definitions are not equivalent in general, the natural question is what is the right definition for a SRB-measure in the case of coupled map lattices. The definitions of naturalness and observability assume that there exists some preferable measure on the space. On finite dimensional manifolds the Lebesgue measure (or some smooth modification of it) has clearly a special role. So, if $X = \prod_{\mathbb{Z}^d} S^1$, one may consider the infinite product of normalized Lebesgue measures on S^1 which is a probability measure on X. However, if one modifies the Lebesgue on S^1 just a little bit and takes the product of these measures, then one obtains a measure which is singular to the product of Lebesgue measures. This indicates that there is no unique preferable measure for coupled map lattices to define naturalness or observability. For the same reason the fourth definition is not directly applicable. But keeping in mind that coupled map lattices are models for real life phenomena, one may argue that, although the real life systems are high dimensional, they are nevertheless finite dimensional. At least one can make

only finite number of observations of finite number of particles. One may also interpret a measurement as a projection so it is natural to consider the finite dimensional projections of an invariant measure. In a finite dimensional subsystem it does not matter whether one uses Lebesgue measure or some modification of it. Therefore Definition 1.1 is a very natural definition.

In statistical physics a phase transition is associated with the non-uniqueness of equilibrium states. (At the temperature 0 degrees of Celsius there are two phases – water and ice.) An interesting question is whether there could be a phase transition in coupled map lattices even for small coupling due to the infinite dimensionality of the system although for the local dynamics there were unique SRB-measure. If one takes as the definition of SRB-measure the one saying that it is an equilibrium state for a potential constructed from the derivative of the map, then the results of [10, 37] imply that there are no phase transitions. If one adopts Definition 1.1, then the results of [33] imply that there is a phase transition even for small coupling. To resolve this paradox, one should analyze the results of [33] more carefully. Indeed, these results imply that for any finite $\Lambda \subset \mathbb{Z}^d$ there is an open set $U \subset X_\Lambda$ with positive Lebesgue measure such that for $\mathcal{L}^\Lambda$-almost all $x \in U$ there are boundary conditions x_1^c and x_2^c such that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \delta_{T^i(x,x_i^c)} = \mu_i, \quad i = 1, 2$$

and $\mu_1 \neq \mu_2$. However, in order to see a measure which is different from the SRB-measure constructed by Bricmont and Kupiainen, one has to choose the boundary condition carefully. In other words, the boundary condition depends on x, and one may argue that the phase transition implied by Theorem 3.1 is not a physical one.

In statistical physics it is natural to talk about equilibrium states associated to interactions since the interaction is the primary concept. However, from the point of view of dynamical systems interaction is rather a tool than a fundamental basic quantity. Thus one would like a more dynamical definition for SRB-measure. Although being natural, Definition 1.1 has the drawback that SRB-measure is not unique for small coupling. I motivated this definition by saying that a measurement is a projection. However, projection means that one takes an average over all boundary conditions. Since one can make only finitely many measurements, one studies also finitely many boundary conditions. Thus it is perhaps better to study conditional distributions on finite dimensional subsystems and demand that they are absolutely continuous with respect to the Lebesgue measure. This approach has been taken by Keller and Zweimüller in [45]. They proved the uniqueness of SRB-measure in this sense. However, they have to assume that the coupling is unidirectional over the one dimensional lattice $\mathbb{N}$ meaning that the boundary conditions have an effect on finite subsystems but the finite subsystems have no effect on the

boundary conditions. Although this assumption is reasonable one would like to relax it.

One possible alternative is to say that μ is a SRB-measure if it is natural or observable with respect to all measures of the form $\prod_{\mathbb{Z}^d} \nu$ where ν is absolutely continuous with respect to the local Lebesgue measure. A step in this direction is taken by Bardet in [5].

As a concluding remark I emphasize that although the theory of SRB-measures for weakly coupled expanding or hyperbolic maps is already quite well understood there are still some very basic problems to be solved.

Acknowledgements

I thank Maarit Järvenpää, Mika Leikas, and Tapani Tolonen for many valuable comments.

References

1. V. Baladi: *Positive transfer operators and decay of correlations*, Advanced Series in Nonlinear Dynamics **16**, (World Scientific Publishing Co. Inc., River Edge NJ 2000) 98
2. V. Baladi: *Periodic orbits and dynamical spectra*, Ergodic Theory Dynam. Systems **18**, (1998), pp 255–292 98
3. V. Baladi, M. Degli Esposti, S. Isola, E. Järvenpää, and A. Kupiainen: *The spectrum of weakly coupled map lattices*, J. Math. Pures Appl. **77**, (1998), pp 539–584 96
4. V. Baladi and H.H. Rugh: *Floquet spectrum of weakly coupled map lattices*, Comm. Math. Phys. **220**, (2001), pp 561–582 96
5. J.-B. Bardet: *Limit theorems for coupled analytic maps*, Probab. Theory Related Fields **124**, (2002), pp 151–177 111
6. M. Blank and L. Bunimovich: *Multicomponent dynamical systems: SRB measures and phase transitions*, Nonlinearity **16**, (2003), pp 387–401 109
7. M. Blank, G. Keller, and C. Liverani: *Ruelle-Perron-Frobenius spectrum for Anosov maps*, Nonlinearity **15**, (2002), pp 1905–1973 96
8. R. Bowen: *Equilibrium states and ergodic theory of Anosov diffeomorphisms*, Springer Lecture Notes in Math. **470**, (1975). 98
9. J. Bricmont and A. Kupiainen: *Coupled analytic maps*, Nonlinearity **8**, (1995), pp 379–396 96, 100, 101
10. J. Bricmont and A. Kupiainen: *High temperature expansions and dynamical systems*, Comm. Math. Phys. **178**, (1996), pp 703–732 96, 98, 101, 110
11. L.A. Bunimovich: *Coupled map lattices: one step forward and two steps back*, Chaos, order and patterns: aspects of nonlinearity—the "gran finale" (Como, 1993), Phys. D **86**, (1995), pp 248–255 95
12. L.A. Bunimovich and Ya.G. Sinai: *Spacetime chaos in coupled map lattices*, Nonlinearity **1**, (1988), pp 491–516 95, 96
13. P. Collet: *Some ergodic properties of maps of the interval*, Dynamical systems (Temuco, 1991/1992), Travaux en Cours **52** (Hermann Paris 1996) pp 55–91 98

14. R.L. Dobrushin: *Gibbsian random fields for lattice systems with pairwise interactions*, Funct. Anal. Appl. **2**, (1968), pp 292–301 100
15. R.L. Dobrushin: *The description of a random field by means of conditional probabilities and conditions on its regularity.* Theory Prob. Appl. **13**, (1968), pp 197–224 100
16. K.J. Falconer: *Hausdorff dimension and the exceptional set of projections*, Mathematika **29**, (1982), pp 109–115 103
17. K.J. Falconer: *Geometry of Fractal Sets*, (Cambridge University Press, Cambridge 1985) 106
18. K.J. Falconer and J.D. Howroyd: *Projection theorems for box and packing dimensions*, Math. Proc. Cambridge Philos. Soc. **119**, (1996), pp 287–295 103
19. K.J. Falconer and J.D. Howroyd: *Packing dimensions of projections and dimension profiles*, Math. Proc. Cambridge Philos. Soc. **121**, (1997), pp 269–286 103
20. K.J. Falconer and P. Mattila: *The packing dimension of projections and sections of measures*, Math. Proc. Cambridge Philos. Soc. **119**, (1996), pp 695–713 103
21. K.J. Falconer and T.C. O'Neil: *Convolutions and the geometry of multifractal measures*, Math. Nachr. **204**, (1999), pp 61–82 103
22. T. Fischer and H.H. Rugh: *Transfer operators for coupled analytic maps*, Ergodic Theory Dynam. Systems **20**, (2000), pp 109–143 96
23. L. Gross: *Decay of correlations in classical lattice models at high temperature*, Comm. Math. Phys. **68**, (1979), pp 9–27 100
24. X. Hu and S.J. Taylor: *Fractal properties of products and projections of measures in* $\mathbb{R}^d$, Math. Proc. Cambridge Philos. Soc. **115**, (1994), pp 527–544 103
25. B.R. Hunt and V.Yu. Kaloshin: *How projections affect the dimension spectrum of fractal measures*, Nonlinearity **10**, (1997), pp 1031–1046 103
26. B.R. Hunt and V.Yu. Kaloshin: *Regularity of embeddings of infinite-dimensional fractal sets into finite-dimensional spaces*, Nonlinearity **12**, (1999), pp 1263–1275 103, 107
27. B.R. Hunt, T. Sauer, and J.A. Yorke: *Prevalence: a translation-invariant "almost every" for infinite dimensional spaces*, Bull. Am. Math. Soc. **27**, (1992), pp 217–238
28. T. Inoue: *Sojourn times in small neighborhoods of indifferent fixed points of one-dimensional dynamical systems*, Ergodic Theory Dynam. Systems **20**, (2000), pp 241–257 109
29. R.B. Israel: *High-temperature analyticity in classical lattice systems*, Comm. Math. Phys. **50**, (1976), pp 245–257 100
30. E. Järvenpää: *A SRB-measure for globally coupled circle maps*, Nonlinearity **10**, (1997), pp 1435–1469 96
31. E. Järvenpää: *A note on weakly coupled expanding maps on compact manifolds*, Ann. Acad. Sci. Fenn. Math. **24**, (1999), pp 511–517 96
32. E. Järvenpää and M. Järvenpää: *Linear mappings and generalized upper spectrum for dimensions*, Nonlinearity **12**, (1999), pp 475–493 103
33. E. Järvenpää and M. Järvenpää: *On the definition of SRB-measures for coupled map lattices*, Comm. Math. Phys. **220**, (2001), pp 1–12 105, 106, 107, 110
34. E. Järvenpää and T. Tolonen: *Natural ergodic measures are not always observable*, preprint 297, http://www.math.jyu.fi/tutkimus/julkaisut.html 109
35. M. Järvenpää: *On the upper Minkowski dimension, the packing dimension, and orthogonal projections*, Ann. Acad. Sci. Fenn. Math. Diss. **99**, (1994), pp 1–34 103
36. M. Jiang: *Equilibrium states for lattice models of hyperbolic type*, Nonlinearity **8**, (1995), pp 631–659 96

37. M. Jiang and Ya.B. Pesin: *Equilibrium measures for coupled map lattices: existence, uniqueness and finite-dimensional approximations*, Comm. Math. Phys. **193**, (1998), pp 675–711 96, 110
38. W. Just: *Globally coupled maps: phase transitions and synchronization*, Phys. D **81**, (1995), pp 317–340 96
39. W. Just: *Bifurcations in globally coupled map lattices*, J. Statist. Phys. **79**, (1995), pp 429–449 96
40. K. Kaneko (ed): *Theory and applications of coupled map lattices*, Nonlinear Sci. Theory Appl. (John Wiley & Sons Ltd., Chichester 1993) 95
41. R. Kaufman: *On Hausdorff dimension of projections*, Mathematika **15**, (1968), pp 153–155. 102, 103
42. R. Kaufman and P. Mattila: *Hausdorff dimension and exceptional sets of linear transformations*, Ann. Acad. Sci. Fenn. Ser. A I Math. **1**, (1975), pp 387–392 103
43. G. Keller: Completely mixing maps without limit measure, Colloq. Math. **100** (2004), no. 1, 73–76. 109
44. G. Keller and M. Künzle: *Transfer operators for coupled map lattices*, Ergodic Theory Dynam. Systems **12**, (1992), pp 297–318 96
45. G. Keller and R. Zweimüller: *Unidirectionally coupled interval maps: between dynamics and statistical mechanics*, Nonlinearity **15**, (2002), pp 1–24 110
46. M. Marstrand: *Some fundamental geometrical properties of plane sets of fractional dimension*, Proc. London Math. Soc. (3) **4**, (1954), pp 257–302 102
47. P. Mattila: *Hausdorff dimension, orthogonal projections and intersections with planes*, Ann. Acad. Sci. Fenn. Math. **1**, (1975), pp 227–244 102
48. P. Mattila: *Hausdorff dimension and capacities of intersection of sets in n-space*, Acta Math. **152**, (1984), pp 77–105 103
49. P. Mattila: *Geometry of Sets and Measures in Euclidean Spaces: Fractals and rectifiability*, (Cambridge University Press, Cambridge 1995) 97, 103, 104, 106
50. P. Mattila: *Hausdorff dimension, projections, and the Fourier transform*, Publ. Math. **48**, (2004), pp 3–48 103
51. M. Misiurewicz and A. Zdunik: Convergence of images of certain measures. In: *Statistical physics and dynamical systems* (Köszeg, 1984), pp 203–219, Progr. Phys. **10**, (Birkhäuser Boston, Boston MA 1985) 109
52. Y. Peres and W. Schlag: *Smoothness of projections, Bernoulli convolutions, and the dimension of exceptions*, Duke Math. J. **102**, (2000), pp 193–251 103, 104, 105
53. Ya.B. Pesin and Ya.G. Sinai: Space-time chaos in chains of weakly interacting hyperbolic mappings. Translated from the Russian by V.E. Nazaĭkinskiĭ. Adv. Soviet Math. **3**, In: *Dynamical systems and statistical mechanics* (Moscow, 1991), pp 165–198, (Amer. Math. Soc., Providence RI 1991) 96
54. D. Ruelle: *A measure associated with Axiom A attractors*, Amer. J. Math. **98**, (1976), pp 619–654 98
55. D. Ruelle: *Thermodynamic formalism. The mathematical structures of classical equilibrium statistical mechanics*, Encyclopedia of Mathematics and its Applications 5 (Addison-Wesley Publishing Co., Reading Mass. 1978) 100
56. T.D. Sauer and J.A. Yorke: *Are the dimensions of a set and its image equal under typical smooth functions?*, Ergodic Theory Dynam. Systems **17**, (1997), pp 941–956 103
57. M. Schmitt: *Spectral theory for nonanalytic coupled map lattices*, Nonlinearity **17**, (2004), pp 671–689 96
58. B. Simon: *The Statistical Mechanics of Lattice Gases, Vol 1*, (Princeton University Press, Princeton New Jersey 1993) 100, 101

59. Ya. Sinai: *Gibbs measures in ergodic theory*, Russian Math. Surveys **27**, (1972), pp 21–69 98
60. V. Volevich: *Construction of an analogue of the Bowen-Ruelle-Sinai measure for a multidimensional lattice of interacting hyperbolic mappings* (Russian), Mat. Sb. **184**, (1993), pp 17–36; translation in Russian Acad. Sci. Sb. Math. **79**, (1994), pp 347–363 96
61. L.-S. Young: *What are SRB measures, and which dynamical systems have them?*, J. Statist. Phys. **108**, (2002), pp 733–754 108, 109

Index

Waves and Oscillations in Networks of Coupled Neurons

B. Ermentrout*

Department of Mathematics, University of Pittsburgh, Pittsburgh, PA 15260
bard@euler.math.pitt.edu

1 Introduction

Neural systems are characterized by the interactions of thousands of individual cells called neurons. Individual neurons vary in their properties with some of them spontaneously active and others active only when given a sufficient perturbation. In this note, I will describe work that has been done on the mathematical analysis of waves and synchronous oscillations in spatially distributed networks of neurons. These classes of behavior are observed both in vivo (that is, in the living brain) and in vitro (isolated networks, such as slices of brain tissue.) We focus on these simple behaviors rather than on the possible computations that networks of neurons can do (such as filtering sensory inputs and producing precise motor output) mainly because they are mathematically tractable. The chapter is organized as follows. First, I will introduce the kinds of equations that are of interest and from these abstract some simplified models. I will consider several different types of connectivity – from "all-to-all" to spatially organized. Typically (although not in every case), each individual neuron is represented by a scalar equation for its dynamics. These individuals can be coupled together directly or indirectly and in spatially discrete or continuous arrays.

Neural models are roughly divided into two main classes: spiking models and firing rate models. We will concentrate on spiking models in this chapter. A comprehensive review of neural network models is given in [8]. In spiking models, we track the firing (or spikes) of individual units, while in firing rate models, we are interested mainly in the average frequency that each unit fires. Individual neurons are highly complex spatially extended nonlinear systems. However, in most models that involve networks of neurons, they are reduced to points in space. The description of these point neurons is itself often very complicated involving dozens of nonlinear differential equations. Again, most

* Supported in part by NIMH and NSF

B. Ermentrout: *Waves and Oscillations in Networks of Coupled Neurons*, Lect. Notes Phys. **671**, 341–357 (2005)
www.springerlink.com

modelers simplify the neurons considerably often reducing them to one- or two-dimensional dynamical systems.

The canonical representation of point neurons is due the Hodgkin and Huxley wherein each cell is considered to be an equivalent electrical circuit with a single capacitor and several resistors in parallel. Thus, the neuron satisfies the differential equation:

$$C\frac{dV}{dt} = I - \sum_k g_k(t)(V - E_k) \tag{1}$$

where I is the experimentally applied current, C is the capacitance of the neuron, E_k are constant reversal potentials, one for each ionic species that travels across the membrane, and $g_k(t)$ are the time-dependent conductances (reciprocal of resistances). The key theoretical point that Hodgkin-Huxley advanced is that $g_k(t)$ are themselves dependent on voltage. That is,

$$g_k(t) = \bar{g}_k m^p h^q$$

where p, q are nonnegative integers (sometimes 0), $\bar{g}_k$ is a nonnegative constant, and $x = (m, h)$ satisfy equations of the form:

$$\tau_x(V)\frac{dx}{dt} = x_\infty(V) - x \ .$$

Without these extra variables, the voltage dynamics is just linear. The original Hodgkin-Huxley model for the squid axon membrane has three of these extra variables, one for the potassium current and two for the sodium current. The simplest model (and one which is commonly used in simulations) is called the leaky integrate-and-fire (LIF) model. There is only one current and it is passive:

$$C\frac{dV}{dt} = I - g_L(V - E_L) \ . \tag{2}$$

However, there is an additional rule that states that if $V(t)$ crosses a proscribed value, V_S, then it is reset to $V_R < V_S$ and the neuron is said to have fired. If I is sufficiently large, the neuron will fire repetitively with a frequency, ω given by:

$$\omega^{-1} = \frac{C}{g_L}\log\frac{I - g_L(V_R - E_L)}{I - g_L(V_S - E_L)} \ .$$

This makes sense only if $I > g_L(V_S - E_L)$. For I large, the frequency is linear with I. Another related model is called the quadratic integrate-and-fire (QIF) model:

$$C\frac{dV}{dt} = I + g_L(V - V_L)(V - V_T)/(V_T - V_L) \ . \tag{3}$$

As with the LIF, when $V(t)$ reaches V_S it is reset to V_R. By choosing $V_S = +\infty$ and $V_R = -\infty$, we recover the dynamics for the normal form of a dynamical

system near a saddle-node on a limit cycle [9, 13, 15]. If $I > g_L(V_T - V_L)/4 \equiv I^*$, then the QIF fires repetitively with frequency:

$$\omega = K(I)\sqrt{(I - I^*)}$$

where $K(I)$ is a complicated expression tending to a constant as $V_S \to +\infty$ and $V_R \to -\infty$. Thus, unlike the LIF model, the QIF model has a square-root dependence of frequency on current for large currents.

Almost all of the neural models that we are concerned with in this chapter fire repetitively when sufficient current is injected into them. There are essentially two mechanisms for going from rest to periodic activity. One of these is the saddle-node on a limit cycle which we have already described. The other mechanism is through a Hopf bifurcation. Near this bifurcation, the behavior of the neuron is similar to its normal form [4]:

$$\frac{dz}{dt} = z(a - b|z|^2)$$

where a, b, z are complex. If we take $a = 1+i$ and $b = 1$ we recover the "radial isochronous clock" (RIC), see [23].

Finally, there are many interesting cases in which the dynamics of the neuron are essentially two-dimensional with V as one dimension and x as the other, where x is one of the auxiliary variables in the Hodgkin-Huxley formalism. For example, suppose that there are only two ionic species: a linear leak and a calcium current. Then the model has the form:

$$C\frac{dV}{dt} = I - g_L(V - E_L) - g_{Ca}m_\infty(V)h(V - E_{Ca}) \tag{4}$$

$$\tau_h(V)\frac{dh}{dt} = h_\infty(V) - h\ . \tag{5}$$

m_∞ is monotonically increasing while h_∞ is a decreasing function of V. For this model, τ_h is sometimes very large, so that it is justifiable to use singular perturbation methods to analyze it. We will do this later in this chapter.

Now that we have briefly described the dynamics of neurons we can ask how to couple them. Neurons communicate in many ways, but the most common mechanism is through chemical synapses. When the presynaptic neuron fires, it produces currents in the post-synaptic neuron. Synapses are modeled similarly to the ionic channels but the auxiliary variables depend on the presynaptic voltage rather than the postsynaptic potential. Thus, a network of neurons is described by equations of the form:

$$C\frac{dV_j}{dt} = I_j - I_{ionic,j} - \sum_k c_{jk}s_k(V_j - E_{R,k}) \tag{6}$$

$$\frac{ds_j}{dt} = \alpha(V_j)(1 - s_j) - s_j/\tau_j \tag{7}$$

where $c_{j,k}$ are nonnegative constants, $E_{R,k}$ is the reversal potential of the kth synapse and $I_{ionic,j}$ are the intrinsic currents for neuron j. When convenient,

we will go to the continuum limit to study spatially organized systems. Thus, our goal in the rest of this chapter is to describe the behavior of coupled systems of neurons.

2 PRC Theory and Coupled Oscillators

We suppose that each neuron is, by itself, an oscillator and that the interactions between neurons are brief and pulsatile. Then, instead of treating the full equations for each oscillator, we suppose that without any external signal, each cell traverses its limit cycle in a regular periodic manner. Thus, we consider each neuron to lie on a one-dimensional circle and satisfy the equation $\theta' = 1$ where θ is the phase of the cycle. Each time θ crosses an integer, we say that the neuron has spiked. Experimental biologists often treat rhythms of unknown mechanism as such black boxes. In order to do some interesting mathematics on this black box, we ask what the effects of brief external perturbations are on the timing of the spikes of our model. Again, this technique has been used for decades to quantify the behavior of biological oscillators [1, 2, 5, 16, 24]. Suppose that the oscillator fires at $t = 0, 1, 2, \ldots$ in absence of any external perturbations. Now suppose at $t = \phi \in [0, 1)$, we apply a brief stimulus. This will cause the oscillator to fire at some time $t = T(\phi)$. In absence of the stimulus, the oscillator fires at $t = 1$ so that the effect of the stimulus has been to advance the timing by an amount $1 - T(\phi)$. If $T > 1$, then a "negative" advance is actually a delay. We call the function, $\Delta(\phi) = 1 - T(\phi)$, the phase-resetting curve or PRC. The PRCs of cortical neurons have been measured by several groups, notably [19, 21].

Figure 1 shows the PRCs from the three simplified models described in the above text. In each of these models, at a time $t = \phi$ the voltage was incremented by an amount a, either positive or negative. For the experimental neuron, the current is held constant to induce repetitive firing and then a brief square pulse of current is superimposed on this background. We note that for small stimuli, the PRC of the RIC is roughly, $\Delta_{RIC}(\phi) = -a \sin 2\pi\phi$ and for the QIF, it is roughly, $\Delta_{QIF}(\phi) = a(1 - \cos 2\pi\phi)$. The PRC for the RIC is very similar to that measured for the flashing rhythm of Pteroptyx malaccae, a Malaysian firefly species known for its dramatic synchronous displays.

Given that we have computed a PRC for a neuron, how can we use this in a mathematical model? In general, the PRC depends in a complicated fashion on the amplitude but for small stimuli, this is almost linear. Consider, first a pair of identical neural oscillators such that each time one fires, the other is adjusted using its PRC. Formally, we can write [14]:

$$\theta_1' = 1 + \Delta(\theta_1)\delta(\theta_2)$$
$$\theta_2' = 1 + \Delta(\theta_2)\delta(\theta_1)\ .$$

Here δ is the Dirac impulse function. This formulation is readily generalized to arbitrary networks of oscillators and by changing the "1" to, say, ω_j we

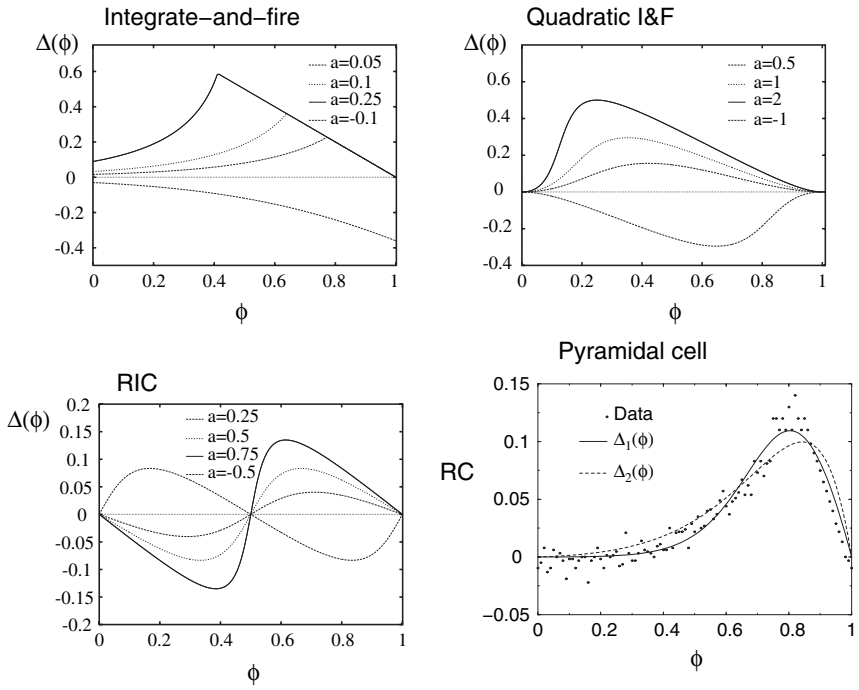


Fig. 1. PRCs from three common models of single neurons and the PRC from a pyramidal cell with some fits to a conductance-based model

can endow each oscillator with its own intrinsic frequency. Let's first study the two-oscillator system by reducing it to a map. Let $F(\phi) = \phi + \Delta(\phi)$. We make two important assumptions about F: (i) $F'(\phi) > 0$ and (ii) $F(0) = 0$. Note that (ii) implies that $F(1) = 1$ since $\Delta(\phi)$ is periodic. Clearly, the PRC for the LIF violates (ii), but the neural PRC and those of the RIC and QIF both satisfy (ii). For weak enough coupling, a small, they also satisfy (i). Physically, these assumptions say that no stimulus can instantly make the neuron fire and that at the moment of spiking, the neuron ignores all stimuli.

We now construct a map for a pair of oscillators. Suppose that when 1 fires, 2 is at ϕ so that the new phase for 2 is $F(\phi)$ and 1 is reset to 0. At $t = t_2 \equiv 1 - F(\phi)$ oscillator 2 fires and oscillator 1 has traveled exactly t_2 since its frequency is 1. Thus, the phase of oscillator 1 is set to $F(t_2)$ and oscillator 2 is set to zero. Finally, oscillator 1 will fire once again at $t_1 = 1 - F(t_2)$ and the phase of oscillator 2 will be t_1. This yields the map

$$\phi \longrightarrow 1 - F(1 - F(\phi)) \equiv G(\phi) \; . \tag{8}$$

We see immediately that $G(0) = 0$ so that there is a synchronous solution. By looking at the iteration, $\phi_{n+1} = G(\phi_n)$, we can analyze the approach to and stability of fixed points. Figure 2 shows $G(\phi)$ for the three oscillators,

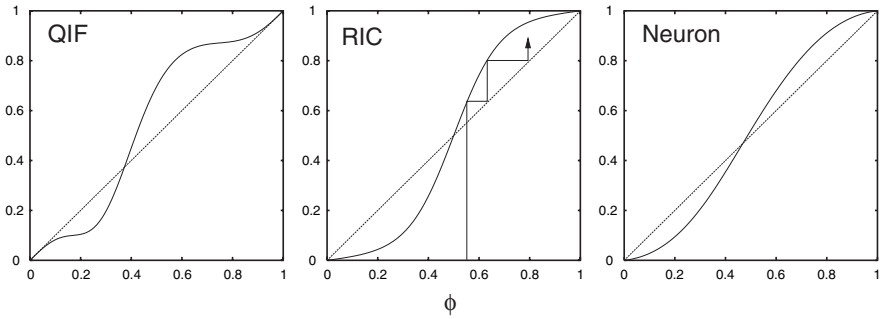


Fig. 2. Maps derived from the PRCs in the previous figure

QIF, RIC, and the real neuron. (For the neuron, we use the approximation, $\Delta(\phi) = \phi(1-\phi)/(1+\exp(-6(\phi-0.6)))$ which is a reasonable fit.) For each of these maps, there is also a non-zero fixed point corresponding roughly to the "anti-phase" solution in which the oscillators fire alternately. The fixed point $\phi = 0$ is stable if and only if $|G'(0)| < 1$ or

$$|F'(1^-)F'(0^+)| = |1 + \Delta'(1^-)||1 + \Delta'(0^+)| < 1 \, .$$

We have used limits, 1^- and 0^+ since the PRCs are continuous, but may not be continuously differentiable at endpoints. (This is certainly true of the neural PRC which is nearly flat at 0 and has a negative slope at 1.) For $a > 0$ in the RIC, synchrony is asymptotically stable as it also is for the neural PRC. The linear stability gives no information for the QIF since $\Delta'(0) = 0$ for this model, but the graphical picture shows that synchrony is stable.

We next turn to some questions about synchrony in more complex networks. Suppose that we have N identical oscillators and each is connected to all the others in exactly the same fashion. Then clearly, synchrony is a solution. In [10] we analyzed this and proved the following condition for stability. Let $\alpha_0 = F'(0^+)$ and $\alpha_1 = F'(1^-)$. Then synchrony is linearly stable if and only if each of the quantities $\beta_l \equiv \alpha_0^l \alpha_1^{N-l}$, $1 \le l < N$ is less than 1. This leads to a rather interesting situation. Suppose that $\alpha_0 > 1$ and $\alpha_1 < 1$ as is the case for the neural PRC. Then for l sufficiently large, we can guarantee that β_l is positive. Thus, for a small network, it is possible that synchrony is stable but for a larger network, it becomes unstable!

Surprisingly, the analysis of synchrony for locally coupled networks is much more difficult than the all-to-all case. The reason for this is that with all-to-all coupling, the ordering of the firing is preserved because of the monotonicity of $F(\phi)$ and the fact that each time an oscillator fires, *all* other oscillators are affected identically. We conjecture that if synchrony is pairwise stable, then it is also stable in a one-dimensional network of nearest-neighbor coupled oscillators. Here, by pairwise stability of synchrony, we mean that for two identical symmetrically coupled oscillators, synchrony will be stable. In one-dimensional rings of sufficient size, there is also the possibility of waves. That

is, there is a constant phase-difference between successive oscillators so that the net phase change around the ring is an integer. Consider the network of N nearest neighbor oscillators:

$$\theta_j' = 1 + \Delta(\theta_j)[\delta(\theta_{j-1}) + \delta(\theta_{j+1})]$$

where we identify 0 with N and $N+1$ with 1. Suppose that when oscillator j fires, oscillator $j-1$ has phase τ so that its new phase is $F(\tau)$. After N iterations of this, we want to have traversed a single cycle. This leads to the following algebraic condition for the existence of a wave:

$$G(\tau, N) \equiv F(F(\tau) + (N-2)\tau) + \tau = 1 \ .$$

Basically, this just says that each oscillator receives two inputs from its two neighbors and that after these it must traverse one cycle. Waves with multiple cycles replace the 1 with m. The wave is linearly stable if and only if (i) $\alpha_N < 1$; (ii) $\alpha_1\alpha_N < 1$; and (iii) $1 + \alpha_1\alpha_N > \alpha_N$, where $\alpha_1 = F'(\tau)$ and $\alpha_N = F'((N-2)\tau + F(\tau))$. The last condition is trivially satisfied if the first two hold. These simple conditions are analogous the conditions for synchrony for a pair of neurons. We point out that this formula is not valid for $\tau = 0$ which is the synchronous solution since it assumes that the firing order is fixed (which is valid near a traveling wave).

In two-dimensions, it is possible to find rotating waves. For example, consider a 4×4 network with nearest neighbor coupling using the RIC PRC. Cells at the corners receive only two inputs, those at the edges, three, and interior cells, four from their neighbors to the north, east, south, and west. The following table summarizes the various firing times in terms of four $((N/2)^2)$ unknowns:

0	α	$\tau/4 - \beta$	τ
$\tau - \beta$	γ	$\tau/4 + \gamma$	$\tau/4 + \alpha$
$3\tau/4 + \alpha$	$3\tau/4 + \gamma$	$\tau/2 + \gamma$	$\tau/2 - \beta$
$3\tau/4$	$3\tau/4 - \beta$	$\tau/2 + \alpha$	$\tau/2$

The structure of this table depends crucially on the fact that the PRC we use is odd-symmetric. Like in the work [18], we can exploit this symmetry to reduce the number of equations. In [18] further symmetries allowed us to reduce the system even more. If $\Delta(-\phi) \neq -\Delta(\phi)$ then there can generally be no reduction. We can use this to derive a series of algebraic conditions for the existence of a rotating wave. Goel and Ermentrout [10] find these explicitly for $N = 4, 6, 8$ and also continue the solutions in the amplitude parameter a for the RIC.

Before moving on to propagation of waves in active media, I want to state a general result about coupled PRCs which is possible when we replace the impulse coupling with a smooth version. Consider the following model:

$$\frac{d\theta_j}{dt} = 1 + \sum_k c_{jk} P(\theta_k)\Delta(\theta_j) \ . \tag{9}$$

The function $P(\theta)$, called the *coupling function*, should be regarded as a pulse-like function; for example $P(\theta) = A\exp(-B(1-\cos 2\pi\theta))$ where B is a large positive number and A is chosen so that the area under P over one period is 1. As $B \to \infty$ this function tends to an impulse function. Suppose that $c_{jk} \geq 0, c_{jk} = c_{kj}$ and that

$$\sum_k c_{jk} = c$$

is a constant. The last assumption implies that $\theta_j(t) = \phi(t)$ where

$$\frac{d\phi}{dt} = 1 + c\Delta(\phi)P(\phi) \equiv f(\phi) \ .$$

We assume that $f(\phi) > 0$ so that $\phi(t)$ is a periodic solution with period:

$$T = \int_0^1 \frac{ds}{f(s)} \ .$$

The first two assumptions allow us to find a simple condition for the stability of the synchronous state. Suppose the matrix c_{jk} is irreducible and

$$Q \equiv \int_0^T P(\phi(t))\Delta'(\phi(t))\ dt < 0 \ .$$

Then the synchronous state is asymptotically stable. The condition on Q is intuitively appealing and has a simple interpretation. Suppose that P is positive near the firing time and nearly zero the rest of the time. Then we require the slope of the PRC to be negative at the firing phase. This is not true for the QIF or the real neuron, but it is true for the RIC. However, for the real neuron, the PRC has a big negative slope on one side of 0 and a very shallow positive slope on the other, so for a symmetric pulse, P, $Q < 0$ as required. The assumptions on c_{jk} are not unreasonable, for example, any ring of oscillators with symmetric positive coupling will work.

3 Waves in Spiking Models

In the first half of these notes, we examined coupled networks of oscillatory neurons. I want to turn now to networks of synaptically coupled units which are not intrinsically oscillatory. Rather, we will assume that the network is "excitable"; that is small perturbations decay to rest, but large enough ones lead to an action potential. Each cell could be a full conductance-based model such as (1), or a simpler scalar model such as the LIF or QIF model. Consider the following general class of scalar models:

$$\frac{dV(x,t)}{dt} = f(V(x,t)) + g\left[V(x,t), \int_\Omega k(x,y)s(y,t)\,dy\right] \tag{10}$$

The functional $s(x,t)$ is effect of the cell at x firing. This could satisfy its own differential equation or be a proscribed function of t or just a nonlinear function of V.

3.1 Theta Model

Let me start with a simple model in which the function s has no independent dynamics. Recall the QIF model (3). This is a scalar model, but it is not convenient for mathematical analysis, so that we transform it to the so-called theta model by setting $V = V_0 + V_1 \tan(\theta/2)$:

$$\frac{d\theta}{dt} = 1 - \cos\theta + (1 + \cos\theta)I \tag{11}$$

where I represents all the inputs. The spiking threshold is $\theta = \pi$ or $V = +\infty$ and the reset is $\theta = -\pi$ or $V = -\infty$. Since $-\pi = \pi$, the theta model is smooth around the spiking so that it represents a smooth dynamical system on the circle. If $I < 0$ then θ tends to a stable fixed point and if $I > 0$, then θ traverses the circle with frequency $\sqrt{I}/\pi$. (See Fig. 3A.) Suppose that the coupling function is as in the previous section; just a pulse like function of θ (see Fig. 3B) and that the domain is the whole line with homogeneous coupling. Then, we arrive at:

$$\frac{d\theta(x,t)}{dt} = 1 - \cos\theta + (1 + \cos\theta)\left(I + g\int_{-\infty}^{\infty} K(x-y)P(\theta(y,t))\,dy\right).$$

$K(x)$ is nonnegative, symmetric, and monotone non-increasing for $x > 0$. $I < 0$ so that the medium is not spontaneously active. Intuitively, we might suspect that if we excite a region above threshold, then the coupling might induce propagation of activity and lead to a traveling wave. This is in fact true under a rather broad range of conditions. Figure 3C shows a simulation of the model and the profiles of two values of $\theta(x,t)$. A traveling pulse satisfies, $\theta(x,t) = U(\xi)$ where $\xi = x - ct$. At $\xi = +\infty$ we want $U(\xi) = \theta_{rest}$ and at $x = -\infty$, $U(\xi) = 2\pi + \theta_{rest}$. This says that the wave traverses exactly one cycle. While there is no published proof of the existence of such a wave front for this model, existence and stability follow with minor changes from a theorem of X. Chen [6] as long as

$$Q(u) = 1 - \cos u + (1 + \cos u)(I + gP(u))$$

has two roots in $[0, 2\pi)$. If $I < 0$, $P(u)$ is narrow enough, and the peak, θ_T of $P(u)$ is close enough to π, then $Q(u)$ will have the required roots. In two spatial dimensions, we can expect spiral waves as shown in part D of the figure.

Osan et al. [17] have considered the same theta model but the coupling function is no longer a simple function of the state θ. Instead, $s(x,t) = \exp[-(t - t^*(x))/\tau]$ where $t^*(x)$ is the time at which $\theta(x,t)$ crosses π. They proved the existence of a traveling wave solution to this problem and in a subsequent paper, Rubin [20] formulated the stability questions.

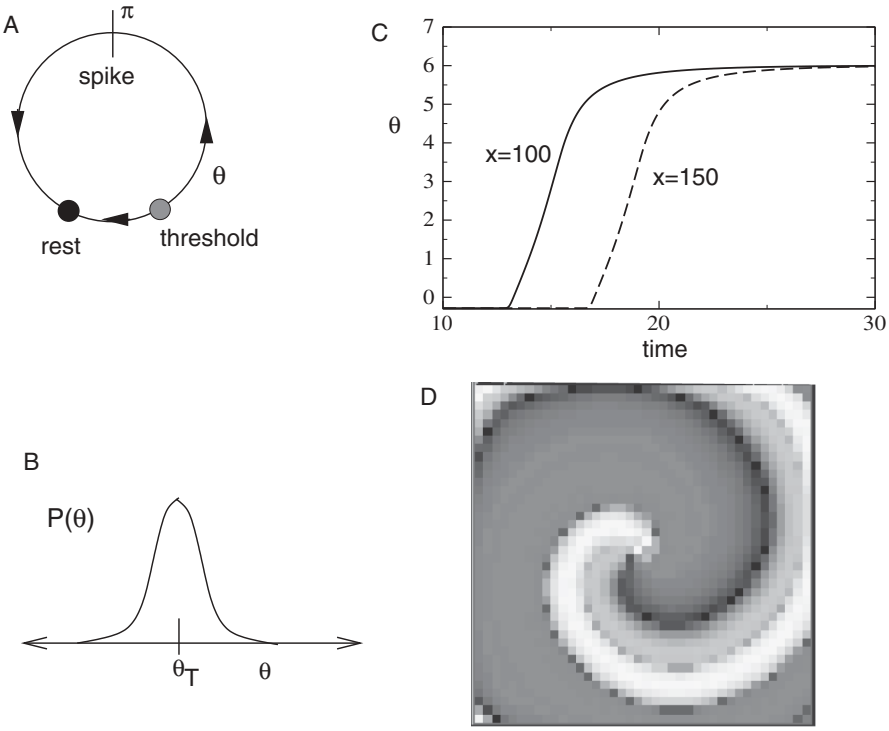


Fig. 3. Waves in a theta model. (**A**) Phase space of the theta model; (**B**) Coupling function of θ; (**C**) Traveling wave for a line of cells; (**D**) Spiral wave for a two-dimensional array

3.2 Thalamic Models

We conclude with a rather complicated conductance-based model of a region of the brain called the thalamus. At its simplest, the network consists of two layers of neurons, the thalamocortical cells (TC) and the reticular nucleus cells (RE). Each neuron satisfies an equation of the form (4) with an additional equation for the synapses. The phaseplane for an individual cell is shown in Fig. 4A. Each cell is endowed with a calcium current which produces rebound excitation. That is, suppose the cell is inhibited for a period of time. This raises the V-nullcline as seen in the figure ($s = \phi$). The equilibrium moves toward the new fixed point. If the inhibition is rapidly removed, the V-nullcline falls back to the original position ($s = 0$) which leaves (V, h) above $h_{\max}$. This causes the voltage to jump to the right branch of the nullcline (a rebound spike), before returning to rest. (More details on this cycle are provided in the next paragraph.) If the two layer network of these cells is wired up as in Fig. 4B, then under some circumstance, the result is a wave of activity across the network. Such a wave is shown in Fig. 4C. This is not a smooth wave;

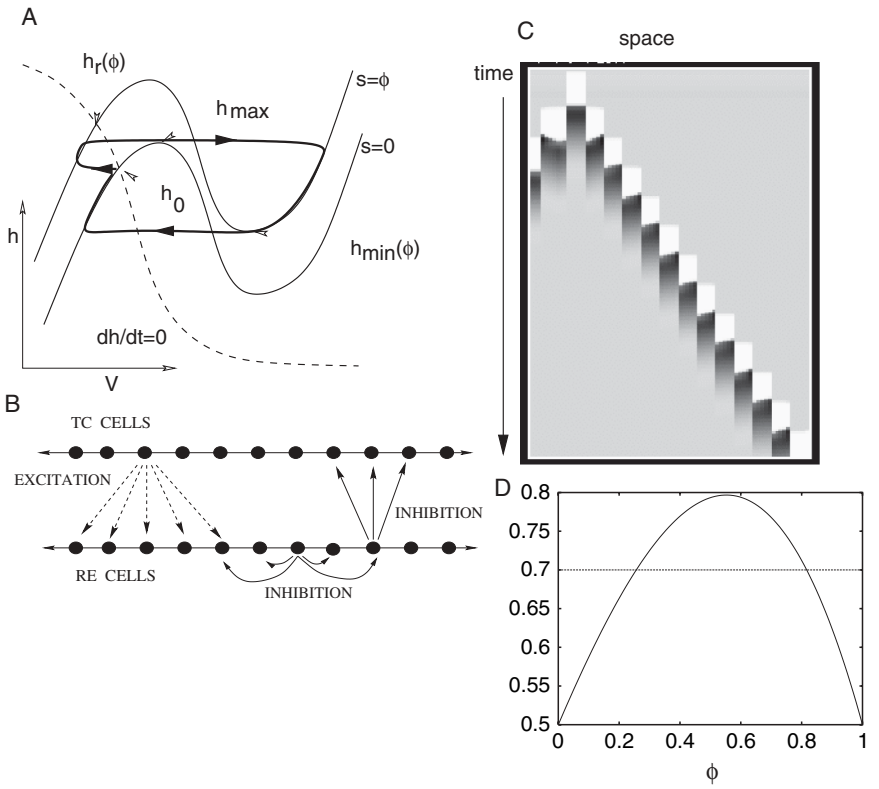


Fig. 4. Thalamic network model. (**A**) Phaseplane showing the h-nullcline (*dashed*) and the V-nullcline at rest ($s = 0$) and when a region of length ϕ is inhibiting ($s = \phi$). Several important values of h are shown. The approximate singular trajectory of a lurching wave is drawn in *thick lines*. (**B**) The architecture of the full model. (**C**) A simulation of a lurching wave. *Grey scale* depicts voltage; white = 40 mv and black = −90 mV. (**D**) The function $F(\phi)$ from (15) with $h_{\max} = .7$, $h_{\min}(\phi) = .2 + .5\phi$, $h_r(\phi) = .5 + \phi$, $\tau_R/\tau_L = 2$

rather we call this a lurching wave. Here is what happens. A group of TC cells fires. This excites RE cells nearby causing them to fire. They inhibit the TC cells including those surrounding the original population of firing cells. The fresh population is inhibited and when the inhibition wears off, they fire as a group.

We now attempt to explain this and find a formula for the size of the groups that fire as well as the time it takes for them to fire. In order to do this, we will simplify a bit and consider a single layer of cells with inhibitory coupling. Thus, when a group of cells fires, it inhibits a neighboring group. After the inhibition wears off, the next group fires and so on. (In the next section, we will look at the transition from smooth waves to lurchers through a different

set of simplified models.) To analyze this model, we use singular perturbation. The present exposition will be a drastic approximation to a fuller analysis of the model which can be found in [22]. The equations of interest are

$$\epsilon \frac{\partial V}{\partial t} = f(V,h) - g \int_{-\infty}^{\infty} W(x-y)s(y,t)\ dy(V - V_{in}) \tag{12}$$

$$\frac{\partial h}{\partial t} = (h_\infty(V) - h)/\tau(V) \tag{13}$$

$$\epsilon \frac{\partial s}{\partial t} = \alpha H(V - V_T)(1-s) - \beta s\ . \tag{14}$$

$f(V,h)$ is all the intrinsic currents of the cell. We suppose that the interaction kernel, $W(x)$ is a square and without generality, assume it is zero outside of $x = (-1/2, 1/2)$. We assume that $\tau(V)$ takes on two values, τ_R when V is on the right branch of the V-nullcline and τ_L when V on the left branch. The parameter ϵ is small indicating that the dynamics is governed by the calcium recovery variable, h. (We should really start with ϵ multiplying the right-hand side of the h equation and then rescale time. In the interest of brevity, I have already done this.) By letting $\epsilon \to 0$, we analyze the singular trajectory and compute the properties of the wave. Figure 4A shows the singular trajectory of a group of cells. Suppose that ϕ is the size of the group of cells that is turned on. This inhibits a neighboring group, which are all at rest, h_0. As long as the first group remains on the right-branch of the V-nullcline, the second group crawls up the left branch of V-nullcline toward $h_r(\phi)$ the resting state. When the first group reaches the bottom of the inhibited ($s = \phi$) V-nullcline, the inhibition disappears. All cells that are above $h_{\max}$ will jump to the right branch, starting the cycle again. With this simple description of the wave, we can derive formulas for the time between jumps and the size of group that jumps. Suppose the group that jumps is on $x \in (-\phi, 0)$. This means that all the synaptic variables, $s(x,t)$, in the group are at their equilibrium values, $a \equiv \alpha/(\alpha+\beta)$. Thus

$$S_{tot}(x) \equiv \int_{-\infty}^{\infty} W(x-y)s(y,t)\ dy = a \int_{-\phi}^{0} W(x-y)\ dy\ .$$

We assume that ϕ is smaller than the synaptic footprint, so that $S_{tot}(x) = a\phi$ for $x \in (0,\phi)$. For the time in which this group of cells in on the right branch of the V-nullcline, all cells in $(-\phi,\phi)$ feel the same common inhibition parameterized by ϕ and see the V-nullcline labeled $s = \phi$ in the figure. At the up-jump, all the cells roughly jump horizontally with $h = h_{\max}$, the maximum of the $s = 0$ V-nullcline. They remain on the right branch until they reach $h_{\min}(\phi)$ where they jump back. For V on the right branch, $h_\infty(V) = 0$, so the time it takes is

$$T = \tau_R \ln \frac{h_{\max}}{h_{\min}(\phi)}\ .$$

In the meantime, the group of cells at $x \in (0, \phi)$ is feeling the inhibition; So, starting from rest, they are heading toward the upper equilibrium point at $h_r(\phi)$. During this period h satisfies

$$\tau_L \frac{dh}{dt} = h_\infty(V_L(h)) - h, \quad h(0) = h_0 \ ,$$

where $V_L(h)$ is the value of the voltage on the $s = \phi$ V-nullcline. We approximate $h_\infty(V_L(h))$ as $h_r(\phi)$ so that we can solve for h in this time period:

$$h(x,t) = h_r(\phi) + (h_0 - h_r(\phi))e^{-t/\tau_L} \ .$$

When the inhibition wears off at $t = T$ only the cells above $h_{\max}$ will make the jump. Thus, for self-consistency, we must have

$$h(\phi, T) = h_{\max}$$

or:

$$h_{\max} = h_r(\phi) + (h_0 - h_r(\phi)) \left(\frac{h_{\min}(\phi)}{h_{\max}} \right)^{\tau_R/\tau_L} \equiv F(\phi) \ . \tag{15}$$

Thus, we reduce the problem to a single equation for ϕ. $F(0) = h_0 < h_{\max}$ so that for small ϕ, $F(\phi) < h_{\max}$. If the synaptic coupling g is large enough and τ_R/τ_L is also large, then $F(\phi)$ will be larger than $h_{\max}$ for ϕ in some range. Indeed, the function $F(\phi)$ is parabolic in shape (see Fig. 4, so there will either two roots or no roots to this equation. Having found ϕ we can then plug it back to get T and thus get the speed of the lurching wave. In [22] a more correct and precise analysis is presented and compares very closely to the solutions obtained by numerically simulating the full model.

3.3 Integrate-and-fire

We close this chapter with an example system which shows the transition from smooth waves to lurching waves as a parameter varies. We will examine the simplest neuronal dynamics, the LIF, in which there is a fixed delay between the time that a cell fires and the time that it excites the neighboring cells. For small delays, the system is like the coupled theta-cell model and a solitary smoothly propagating wave exists and is stable. For large delays, we can think of this as analogous to the thalamic model based on inhibition and rebound. The delay is analogous to the time for the inhibition to wear off allowing the stimulated cell to fire; if this is large enough lurching occurs. Suppose that each time a cell fires, the result on a neighboring cell is a proscribed function of time, $\alpha(t)$. The domain is the real line. Then the equation of each cell is:

$$\tau_0 \frac{\partial V(x,t)}{\partial t} = -V(x,t) + g_{syn} \int_{-\infty}^{\infty} w(x-y)\alpha(t - T(y) - \tau_d) \ dy + I \ ,$$

where I is a constant applied current and $T(x)$ is the time at which the cell at x fires. We say that a cell fires when $V(x,t)$ crosses V_T, the threshold. Typically, $\alpha(t) = (\exp(-t/\tau_1) - \exp(-t/\tau_2)/(\tau_1 - \tau_2)$ for $t > 0$ and 0 otherwise. $w(x)$ will be either a square, Gaussian, or exponential interaction.

Since V satisfies a linear differential equation, we can integrate this equation once to obtain a Volterra integral equation for $V(x,t)$. We then note that $V(x,T(x)) = V_T$ by definition which leads to a very interesting functional equation for $T(x)$:

$$\frac{V_T}{g_{syn}} = \int_{-\infty}^{\infty} dy\ w(y)G[T(x) - T(y) - \tau_d]\ , \tag{16}$$

where

$$G(t) = \int_0^{\infty} e^{-(t-s)/\tau_0}\alpha(s)\ ds\ .$$

Equation (16) is equivalent to the original model under the assumption that each cell fires at most one time. We could justify the "one-spike" assumption by assuming a very strong hyperpolarizing afterpotential or strong synaptic depression, both preventing further spikes.

Figure 5A shows the firing pattern of a network of LIF neurons with two different delays. In the left, a smooth wave appears so that $T(x) = x/\nu$ where ν is the wave velocity. On the right, this simple description no longer holds. However, we do have that $T(x+L) = T(x) + T_{per}$ and the average velocity is $\nu = L/T_{per}$. That is, $T(x) - x/\nu$ is a periodic function of x. The smooth wave is simple to analyze since we know the exact form of $T(x) = x/\nu$. Plugging this into (16), we obtain:

$$\frac{V_T}{g_{syn}} = \int_0^{\infty} dy w(y + \tau_d\nu)G(y/\nu) \equiv F(\nu)\ .$$

The function $F(\nu)$ can be analytically determined in several interesting cases, for example if $w(x) = \exp(-|x/\sigma|)/(2\sigma)$ and $\tau_1 = 0$, we find

$$F(\nu) = \frac{2(\tau_0\nu + \sigma)(\tau_2\nu + \sigma)}{\tau_0\nu\sigma}\exp(\tau_d\nu/\sigma)\ .$$

In general, this is a parabolic function of ν so that there are either no roots, one root, or two roots. Figure 5B shows the velocity as a function of the threshold for several different parameters and delays. It was showed in [7] showed that this was generally the case for conductance-based neurons using a combination of numerical shooting and asymptotics.

Following [12] or [3], we can look at the stability of these traveling waves. Substituting, $T(x) = x/\nu + \epsilon e^{\lambda x}$ into (16) and taking the $O(\epsilon)$ terms, we obtain an equation for λ:

$$E(\lambda) \equiv \int_0^{\infty} dy w(y + \tau_d\nu)G'(y/\nu)\left[1 - e^{-\lambda(y+\tau_d\nu)}\right] = 0\ . \tag{17}$$

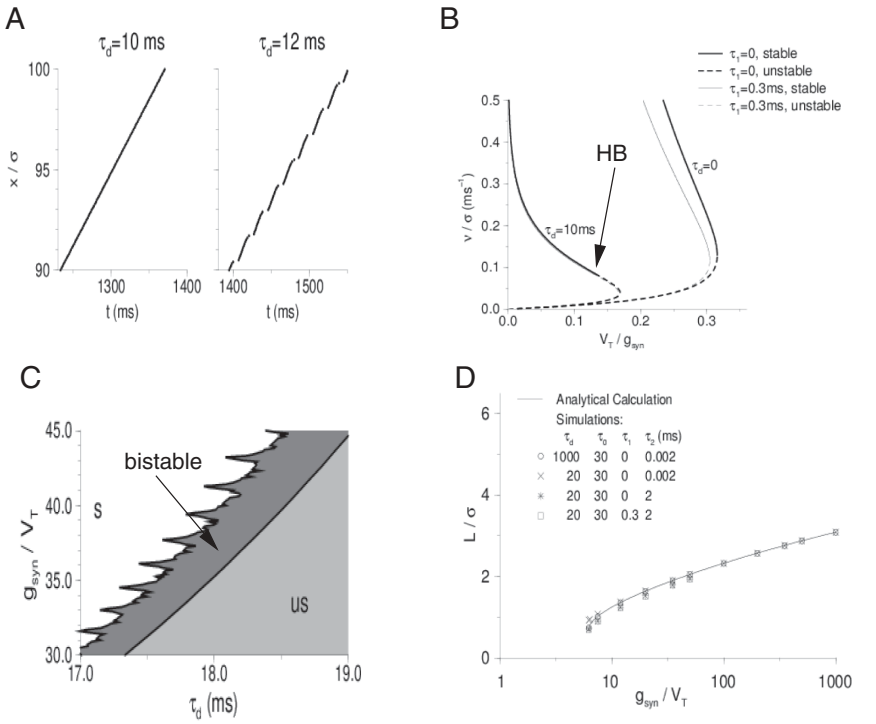


Fig. 5. The integrate-and-fire model with delays. (**A**) a raster plot showing the time of firing of cells in a numerically computed wave for two delays. (**B**) velocity as a function of the threshold for different α functions. Note the Hopf bifurcation leading to instability of the smooth wave. (**C**) Complex behavior of the network; white region – only smooth waves are stable; *light grey* – only lurching waves are stable; *dark grey* – both waves are stable. (**D**) Lurching period as a function of synaptic strength

The function $E(\lambda)$ is called the Evans function; $E(0) = 0$ since there is translation invariance. If the real part of λ is positive, the perturbation of the wave will grow exponentially as it advances through space and it is thus unstable. It was showed in [3] and [12] that the lower branch of solutions (the slow waves) are unstable for all parameters. The fast waves are stable if there is no delay. What is relevant to lurching waves is that Golomb and Ermentrout showed in [12] that for a large enough delay, the fast waves lose stability at a Hopf bifurcation and give rise to waves that have periodic modulation. That is, $T(x) - x/\nu$ is a periodic function. We have not yet computed the direction of bifurcation for this problem; however, we have done careful numerical analysis the results of which are shown in Fig. 5C. There is a complex regime of bistability between lurching waves and smooth waves. We conjecture that the borders of this region are precisely those places where the Hopf bifurcation

switches from sub- to super-critical. (Oscillations branching from subcritical Hopf bifurcations undergo turning points to become stable leading to a regime of bistability between the fixed point and the oscillation.) These results suggest that the transition between smooth waves and lurching waves in the thalamic model is a consequence of changing the effective delay to excitation. Golomb and Ermentrout showed [11] simulations of a conductance-based model of TC and RE cells that undergoes a transition from smooth to lurching waves as the strength of inhibition (and thus the effective delay to firing) increases.

As in the previous section of this chapter, we can estimate the size of a lurching group of neurons in certain asymptotic regimes. Recall that for the singular perturbation arguments, we let ϵ, the recovery rate for the T-current tend to zero. In unscaled time, this means that the period between lurching groups tends to infinity. The analogue in the present model is to let the delay become arbitrarily large. We obtained the following estimate for the size, L of a group of neurons which fire together during a lurching cycle:

$$\frac{V_T}{g_{syn}} = \int_L^{2L} dy\ w(y) \equiv W(L)\ .$$

This implicitly defined function gives an excellent approximation for the size of the lurching group as shown in Fig. 5D. For an exponential footprint,

$$L \sim \sigma \ln\left(\frac{g_{syn}}{2V_T}\right)$$

for large g_{syn} while for a Gaussian footprint,

$$L \sim \sqrt{2}\sigma\sqrt{\ln(g_{syn}/V_T)}\ .$$

References

1. J. Ayers, A.I. Selverston, *Synaptic perturbation and entrainment of gastric mill rhythm of the spiny lobster*, J. Neurophysiol. **51** (1984), no. 1, 113–25. 344
2. E.N. Best, *Null space in the Hodgkin-Huxley Equations. A critical test*, Biophys. J. **27** (1979), no.1, 87–104. 344
3. P.C. Bressloff, *Traveling waves and pulses in a one-dimensional network of excitable integrate-and-fire neurons*, J. Math. Biol. **40** (2000), no. 2, 169–198. 354, 355
4. E. Brown, J. Moehlis, and P. Holmes. *On the phase reduction and response dynamics of neural oscillator populations*, Neural Computation **16** (2004), 673–715. 343
5. C.C. Canavier, R.J. Butera, R.O. Dror, D.A. Baxter, J.W. Clark, J.H. Byrne, *Phase response characteristics of model neurons determine which patterns are expressed in a ring circuit model of gait generation*, Biol. Cybern. **77** (1997), no. 6, 367–80. 344
6. X. Chen, *Existence, uniqueness, and asymptotic stability of traveling waves in nonlocal evolution equations*, Adv. Differential Equations **2** (1997), no. 1, 125–160. 349

7. B. Ermentrout, *The analysis of synaptically generated traveling waves*, J. Computat. Neurosci. **5** (1998), 191–208. 354
8. B. Ermentrout, *Neural networks as spatio-temporal pattern-forming systems*, Reports on Progress in Physics **61** (1998), 353–430. 341
9. B. Ermentrout, N. Kopell, *Parabolic bursting in an excitable system coupled with a slow oscillation*, SIAM J. Appl. Math. **46** (1986), no. 2, 233-253. 343
10. P. Goel, B. Ermentrout, *Synchrony, stability, and firing patterns in pulse-coupled oscillators*, Physica D **163** (2002), no. 3-4, 191–216. 346, 347
11. D. Golomb, B. Ermentrout, *Continuous and lurching waves in a neuronal network with delay and spatially decaying connectivity*, Proc. Nat. Acad. Sci. U.S.A. **99**:13480-13485 (1999). 356
12. D. Golomb, B. Ermentrout, *Effects of delay on the type and velocity of travelling pulses in neuronal networks with spatially decaying connectivity*, Network Comp. Neural **11** (2000), no. 3, 221–246. 354, 355
13. F.C. Hoppensteadt, E.M. Izhikevich, *Weakly Connected Neural Networks*, Springer-Verlag, New York, 1997. 343
14. E.M. Izhikevich, *Class 1 neural excitability, conventional synapses, weakly connected networks, and mathematical foundations of pulse-coupled models*, IEEE Trans. on Neural Networks **10** (1999), 499–507. 344
15. E.M. Izhikevich, *Neural excitability, spiking, and bursting*, Int. J. of Bifurcation and Chaos **10** (2000), 1171–1266. 343
16. S.A. Oprisan, A.A. Prinz and C.C. Canavier, *Phase resetting and phase locking in hybrid circuits of one model and one biological neuron*, Biophys J. **87** (2004), no. 4, 2283–98. 344
17. R. Osan, J. Rubin, and B. Ermentrout, *Regular traveling waves in a one-dimensional network of theta neurons*, SIAM J. Appl. Math. **62** (2002), 1197–1221. 349
18. J.E. Paullet, B. Ermentrout, *Stable rotating waves in two-dimensional discrete active media*, SIAM J. Appl. Math. **54** (1994), 1720–1744. 347
19. A.D. Reyes, E.E. Fetz, *Effects of transient depolarizing potentials on the firing rate of cat neocortical neurons*, J. Neurophysiol. **69** (1993) 1673 1683. 344
20. J. Rubin, *A nonlocal eigenvalue problem for the stability of a traveling wave in a neuronal medium*, Disc. Cont. Dyn. Sys. A **4** (2004), 925–940. 349
21. R. Stoop, K. Schindler and L.A. Bunimovich, *Neocortical networks of pyramidal neurons: from local locking and chaos to macroscopic chaos and synchronization*, Nonlinearity **13** (2000), 1515–1529. 344
22. D.H. Terman, B. Ermentrout, A.C. Yew, *Propagating activity patterns in thalamic neuronal networks*, SIAM J. Appl. Math. **61** (2001), no. 5, 1578–1604. 352, 353
23. A.T. Winfree, *The Geometry of Biological Time*, Springer, New York, 1980. 343
24. D.L. Ypey, W.P. Van Meerwijk, G. de Bruin, *Suppression of pacemaker activity by rapid repetitive phase delay*, Biol. Cybern. **45** (1982), no. 3, 187–94. 344

Modeling the Dynamics of Genetic Regulatory Networks: Continuous and Discrete Approaches

H. de Jong[1] and R. Lima[2]

[1] Institut National de Recherche en Informatique et en Automatique (INRIA), Unité de recherche Rhône-Alpes, 655 avenue de l'Europe, Montbonnot, 38334 Saint Ismier Cedex
Hidde.de-Jong@inrialpes.fr

[2] Centre de Physique Théorique, CNRS, Luminy, case 907, 13288 Marseille Cedex 09
lima@cpt.univ-mrs.fr

1 Introduction

A remarkable development in molecular biology today is the upscaling to the genomic level of its experimental methods. Hardly imaginable only 20 years ago, the sequencing of complete genomes has become a routine job, highly automated and executed in a quasi-industrial environment. The miniaturization of techniques for the hybridization of labeled nucleic acids in solution to DNA molecules attached to a surface has given rise to DNA microarrays, tools for measuring the level of gene expression in a massively parallel way [1]. The development of proteomic methods based on two-dimensional gel electrophoresis, mass spectrometry, and the double-hybrid system allows the identification of proteins and their interactions at a genomic scale [2].

These novel methods in genomics produce enormous quantities of data about different aspects of the cell. On the one hand, they allow the identification of interactions between the genes of an organism, its proteins, metabolites, and other small molecules, thus mapping the structure of interaction networks. On the other hand, they are able to detect the evolution of the state of the cell, that is, the temporal variation of the concentration and the localization of the different molecular components, in response to changes in the environment. The big challenge of *systems biology* consists in relating these structural and functional data to each other, in order to arrive at a global interpretation of the functioning of the organism [5, 6]. This amounts to predicting and understanding how the observed behavior of the organism – the adaptation to its environment, the differentiation of its cells during

H. de Jong and R. Lima: *Modeling the Dynamics of Genetic Regulatory Networks: Continuous and Discrete Approaches*, Lect. Notes Phys. **671**, 307–340 (2005)
www.springerlink.com

development, even its evolution on a longer time-scale – emerges from the set of molecular interactions.

In addition to high-throughput experimental methods, mathematical and computational approaches are indispensable for the analysis of genetic regulatory networks. Given the large number of components of most networks of biological interest, connected by positive and negative feedback loops, an intuitive comprehension of the dynamics of the system is often difficult, if not impossible to obtain. *mathematical modeling* supported by *computer tools* can contribute to the analysis of a regulatory network by allowing the biologist to focus on a restricted number of plausible hypotheses. The formulation of a mathematical model requires an explicit and non-ambiguous description of the hypotheses being made on the regulatory mechanisms under study. Furthermore, its simulation by means of the model yields predictions on the behavior of the cell that can be verified experimentally.

The use of mathematical methods for the modeling and simulation of molecular interaction networks is not new, of course. The first modeling studies of gene regulation, such as the well-known Goodwin model of genetic autoregulation [10], can be traced back to the very beginnings of molecular biology, while the modeling of metabolic pathways has an even longer pedigree. Nevertheless, it is fair to say that mathematical modeling has not yet been fully integrated in the actual practice of biological research, at least not to the extent this has been the case for other scientific disciplines, most notably physics. This raises the question whether some of the accumulated experience in physics may be transferable to biology, in particular the combined mathematical and experimental analysis of carefully designed and tightly controlled artificial systems. The interest of such systems derives from the fact that they allow the dynamics of basic mechanisms found in every natural system to be studied in isolation.

In this chapter, we will discuss methods for the modeling and simulation of one particular type of molecular interaction network, called *genetic regulatory network*. Today, a large part of the experimental data available, notably gene transcription data, concerns these networks of genes, proteins, and their mutual interactions. Many reviews of the modeling and simulation of genetic regulatory networks have been published in recent years (e.g., [13, 14, 15, 16, 61]), presenting the wide variety of formalisms that have been proposed in the literature, such as oriented graphs, Bayesian networks, Boolean networks, differential equations, and stochastic master equations. We will restrict the discussion here to *ordinary differential equations* and to *iterations of maps*. These continuous-time and discrete-time models have been often used for the modeling of biological systems and a large number of powerful techniques for their analysis are available.

In Sect. 2, we introduce some biological notions, fundamental for understanding the nature of genetic regulation and genetic regulatory networks. Sections 3 and 4 discuss the analysis of a simple example network, a cross-inhibition network of two genes, by means of several continuous-time

differential equations and discrete-time dynamical systems. In addition, references are made to modeling studies of real genetic regulatory networks underlying the functioning and development of several prokaryote and eukaryote systems. The chapter ends with a brief discussion and conclusions.

2 Genetic Regulatory Networks

Genes are segments of DNA involved in the production of proteins, which play a primary role in the chemistry of a living cell [9]. The process in which a protein is produced from the information encoded in the nucleotide sequence of the DNA is called *gene expression*. Gene expression comprises several steps, two of which are shown in Fig. 1. First, the coding region of a gene is transcribed into mRNA by an enzyme called RNA polymerase. The resulting mRNA molecule then functions as a template for the synthesis of a protein by another enzyme, the ribosome, in a process called translation. In eukaryotic organisms, gene expression is considerably more complicated and may involve intermediary steps in which RNA is processed and transported from the cell nucleus to the cytoplasm. The synthesis of proteins is balanced by their degradation, that is, by processes that break down proteins into their amino acid components.

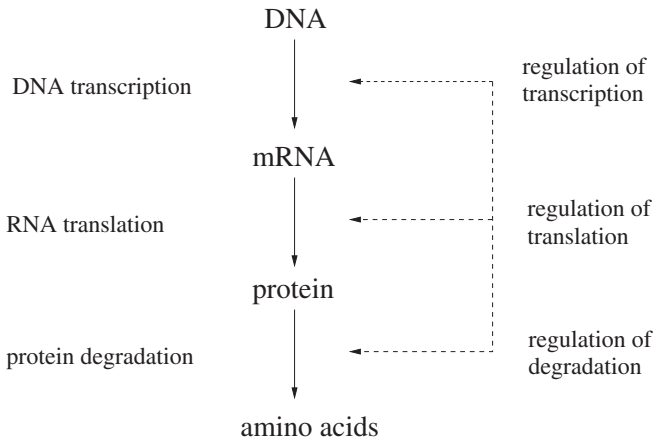


Fig. 1. Regulation of production and degradation of proteins

The level of gene expression, that is, the cellular concentration of the protein encoded by a gene, depends on the relative activity of protein synthesis and degradation. In order to adapt the gene expression level to the requirements of the cell at any given time, complicated mechanisms regulating the production and destruction of proteins have emerged in the course of evolution (Fig. 1). A simple example is the control of transcription by a repressor

protein binding to a regulatory site on the DNA, thus preventing RNA polymerase from transcribing the gene. The example illustrates that the regulation of gene expression involves proteins encoded by other genes. This gives rise to *genetic regulatory systems* structured by networks of interactions between genes, proteins, and other molecules, so-called *genetic regulatory networks*.

We will illustrate the main interactions in a genetic regulatory network by means of a simple example: the regulation of the expression of the sigma factor σ^S in *Escherichia coli*. of the sigma subunits of the RNA polymerase is to recognize specific transcription initiation sites on the DNA, the so-called *promoters*. The expression or activation of a certain sigma factor therefore leads to the expression of a specific subset of genes of the organism. This type of regulation is often used by bacteria to assure a global response to an important change in their environment. Because of their importance for the global functioning of the cell, the expression of the sigma factors themselves is often tightly regulated.

E. coli possesses seven different sigma factors [17, 18]. The principal sigma factor, σ^{70}, directs the transcription of the so-called housekeeping genes. In many stress situations (lack of nutriments, high osmolarity, change of pH or of temperature, etc.), *E. coli* expresses the alternative sigma factor σ^S, encoded by the gene *rpoS*. σ^S takes its name from the fact that it plays an important role in the adaptation to a particular stress, frequently encountered by bacteria: the depletion of nutriments in the environment, which leads to a considerable slowing down of cell growth, called the stationary growth phase. However, σ^S is activated in response to many other kinds of stress [19]. The regulation of the concentration of σ^S in the cell is extraordinarily complex and provides a good illustration of the different modes of regulation shown in Fig. 2.

Although regulation of *transcription* constitutes the preferred mode of regulating gene expression in bacteria, few studies have addressed this subject in the case of *rpoS*. As a consequence, our knowledge on the transcriptional regulation of this gene remains incomplete. The protein CRP, a typical repressor-activator, specifically binds the DNA at two sites close to the major promoter of *rpoS* [19]. One of these sites overlaps with the promoter, which implies that CRP and RNA polymerase cannot simultaneously bind to the DNA, due to sterical constraints. As a consequence, CRP represses the transcription of *rpoS*. The second binding site of CRP is located just upstream of the promoter. This geometry is reminiscent of the *lac* operon, where CRP binding to a similarly-positioned site establishes protein-protein interactions with the RNA polymerase, thus favoring the recruitment of RNA polymerase to the promoter [20]. The molecular details of this apparently contradictory regulation of the transcription of *rpoS* by CRP are still only partially understood. Nevertheless, the example illustrates one type of regulation that is quite widespread in bacteria: a protein binding the DNA (the regulator) prevents or favors the binding of RNA polymerase to the promoter.

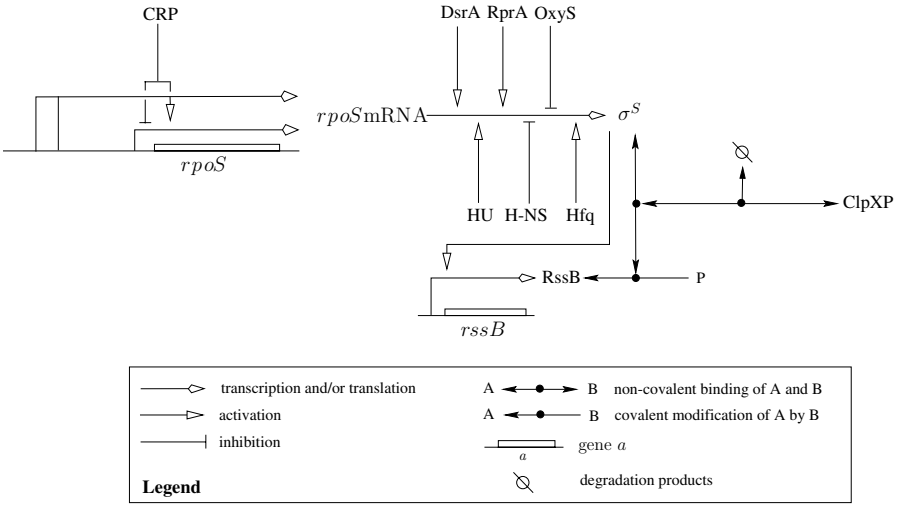


Fig. 2. Regulation of the synthesis and degradation of the sigma factor σ^S. Only the interactions detailed in the text are shown. The notation follows the graphical conventions proposed by Kohn [21]

The expression of *rpoS* is not only regulated at the transcriptional, but also at the post-transcriptional level. The *translation* of the mRNA of *rpoS* is stimulated by environmental stress factors such as high osmolarity, low temperature, or low pH. The translation begins with the recognition of the so-called *Shine-Dalgarno sequence* by the ribosome, followed by the binding of the latter to this sequence. The efficiency of translation depends on the similarity of the Shine-Dalgarno sequence to the consensus sequence, and its accessibility for the ribosome. The mRNA of *rpoS*, like any other RNA, is not only a linear molecule, but possesses a particular secondary and tertiary structure. If the Shine-Dalgarno sequence is sequestered in a secondary structure (e.g., an RNA helix), it will be less accessible to the ribosome and the efficiency of translation will be reduced. In the case of *rpoS*, at least three small regulatory RNAs (DsrA, RprA, OxyS) and an equal number of proteins (HU, H-NS, Hfq) are known to modify the structure of the RNA near the Shine-Dalgarno sequence [19, 22].

Not only the synthesis of σ^S is tightly regulated, but its *degradation* is also subject to multiple environmental and physiological influences [19]. Like many proteins in *E. coli*, σ^S is degraded by proteases. All proteins are recognized by proteases when they are misfolded or truncated. In addition, certain proteins contain sequences (often at the N or C-terminal region) that are specifically recognized by proteases. In the case of σ^S, a highly specialized system targets the protein for degradation by the ATP-dependent protease ClpXP. The protein RssB, when phosphorylated, forms a tight complex with σ^S. RssB also interacts with ClpX, the subunit of the ClpXP complex that recognizes the

substrates of the protease, and thus targets σ^S towards ClpXP. The catalytic subunit, ClpP, degrades σ^S and RssB∼P is released, ready to dispatch another σ^S molecule towards degradation. The system is finely regulated by a feedback loop: the synthesis of RssB depends on σ^S. In addition, the σ^S-RssB system is subject to environmental signals, since RssB only binds to σ^S, if it is phosphorylated in response to a thus far unidentified signal.

Transcription, translation, and degradation are regulated by a large number of interactions, as illustrated here for σ^S. This convergence or "fan-in" of influences on *rpoS* and its protein is accompanied by an even more important divergence or "fan-out", in the sense that σ^S regulates the transcription of at least 70 genes of *E. coli* [19]. Among these genes, several encode proteins involved, directly or indirectly involved in the regulation of the synthesis and degradation of σ^S. This endows the network with a complex feedback structure, responsible for the adaptation of the transcriptional program of the bacterium to external perturbations.

The complexity of the genetic regulatory network is further increased by the fact that it is integrated with other networks. As we mentioned above, by citing the example of RssB, the σ^S regulon is the target of environmental signals. It is also the target of regulatory factors originating in the cellular metabolism. For instance, the expression of σ^S, or at least of genes dependent on σ^S, is also sensitive to the concentration of lactic acid or the redox state of the cell (as measured by the ratio of NADH and NAD^+) [19]. The genetic regulatory network controlling the expression of *rpoS* as well as the regulation of expression of target genes of σ^S is thus embedded in the metabolic and signal transduction networks of the cell. A complete understanding of the dynamics of this system would require a detailed description of all these elements. However, we can often abstract from the metabolic and signal transduction networks – by focusing on their effects on gene expression – and nevertheless obtain an adequate description of the global functioning of the regulatory system [23].

3 Ordinary Differential Equation Models

3.1 Nonlinear Models

Nonlinear ordinary differential equations are probably the most-widespread formalism for modeling genetic regulatory networks. They represent the concentration of gene products – mRNAs or proteins – by continuous, time-dependent variables, that is, $x(t)$, $t \in T$, T being a closed time-interval ($T \subseteq \mathbb{R}_{\geq 0}$). The variables take their values from the set of non-negative real numbers ($x : T \to \mathbb{R}_{\geq 0}$), reflecting the constraint that a concentration cannot be negative. In order to model the regulatory interactions between genes, functional or differential relations are used.

More precisely, gene regulation is modeled by a system of ordinary differential equations having the following form:

$$\frac{dx_i}{dt} = f_i(\boldsymbol{x}),\ i \in [1, \ldots, n]\ , \tag{1}$$

where $\boldsymbol{x} = [x_1, \ldots, x_n]'$ represents the vector of concentration variables of the system, and the function $f_i : \mathbb{R}^n_{\geq 0} \to \mathbb{R}$, usually highly *nonlinear*, the regulatory interactions. The system of (1) describes how the temporal derivative of the concentration variables depends on the values of the concentration variables themselves. In order to simplify the notation, we can write (1) as the vector equation

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{f}(\boldsymbol{x})\ , \tag{2}$$

with $\boldsymbol{f} = [f_1, \ldots, f_n]'$. Several variants of (2) can be imagined. For instance, by taking into account the input variables $\boldsymbol{u}$, it becomes possible to express the dependence of the temporal derivative on external factors, such as the presence of nutriments. In order to account for the delays resulting from the time it takes to complete transcription, translation, and the other stages of the synthesis and the transport of proteins, (2) has to be changed into a system of delay differential equations [16].

The above definitions can be illustrated by means of a simple network of two genes (Fig. 3). Each of the genes encodes a regulatory protein that inhibits the expression of the other gene, by binding to a site overlapping the promoter of the gene. Simple as it is, this *mutual-inhibition network* is a basic component of more complex, real networks and allows the analysis of some characteristic aspects of cellular differentiation [24, 26].

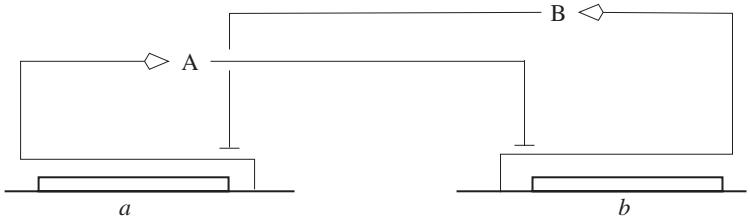


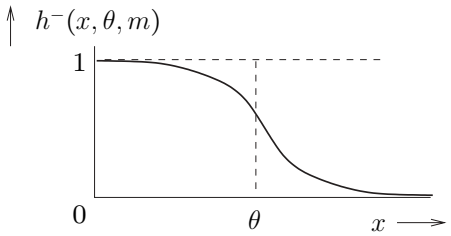
Fig. 3. Example of a simple genetic regulatory network, composed of two genes (**a**) and (**b**), the proteins A and B, and their regulatory interactions

An ordinary differential equation model of the network is shown in Fig. 4(a). The variables x_a and x_b represent the concentration of the proteins A and B, encoded by the genes a and b, respectively. The temporal derivative of x_a is the difference between the *synthesis term* $\kappa_a\, h^-(x_b, \theta_b, m_b)$ and the *degradation term* $\gamma_a\, x_a$. The first term expresses that the rate of

synthesis of protein A depends on the concentration of protein B and is described by the function $h^- : \mathbb{R}_{\geq 0} \times \mathbb{R}^2_{>0} \to \mathbb{R}_{\geq 0}$. This so-called *Hill function* is monotonically decreasing. It takes the value 1 for $x_b = 0$, and asymptotically reaches 0 for $x_b \to \infty$. It is characterized by a threshold parameter θ_b and a cooperativity parameter m_b (Fig. 4(b)). For $m_b > 1$, the Hill function has a sigmoidal form that is often observed experimentally [28, 29]. The synthesis term $\kappa_a\, h^-(x_b, \theta_b, m_b)$ thus means that, for low concentrations of the protein B, gene a is expressed at a rate close to its maximum rate κ_a ($\kappa_a > 0$), whereas for high concentrations of B, the expression of the gene is almost completely repressed. The second term of the differential equation, the degradation term, expresses that the degradation rate of the protein A is proportional to its own concentration x_a, γ_a being a degradation parameter ($\gamma_a > 0$). In other words, the degradation of the protein is not regulated in this example. The differential equation for x_b has an analogous interpretation.

$$\frac{dx_a}{dt} = \kappa_a\, h^-(x_b, \theta_b, m_b) - \gamma_a\, x_a$$
$$\frac{dx_b}{dt} = \kappa_b\, h^-(x_a, \theta_a, m_a) - \gamma_b\, x_b$$
$$h^-(x, \theta, m) = \frac{\theta^m}{x^m + \theta^m}$$

(a)



(b)

Fig. 4. (**a**) Nonlinear ordinary differential equation model of the mutual-inhibition network (Fig. 3). The variables x_a and x_b correspond to the concentrations of proteins A and B, respectively, the parameters κ_a and κ_b to the synthesis rates of the proteins, the parameters γ_a and γ_b to the degradation rates, the parameters θ_a and θ_b to the threshold concentrations, and the parameters m_a and m_b to the degree of cooperativity of the interactions. All parameters are positive. (**b**) Graphical representation of the characteristic sigmoidal form, for $m > 1$, of the Hill function $h^-(x, \theta, m)$

Because of the nonlinearity of the functions $\boldsymbol{f}$, the solutions of the system of ordinary differential equations (2) cannot generally be determined by analytical means. This is even true for the nonlinear model of the two-gene network (Fig. 4). However, because the model has only two variables, we can obtain a qualitative understanding of the dynamics of the network, by applying the tools available for analysis in the phase plane (see [30] for an accessible introduction).

The *phase plane* of the system is represented in Fig. 5. Every point in the plane represents a pair of concentrations x_a and x_b. The solutions of the system of differential equations give rise to *trajectories* in the phase plane, as illustrated in (a). Another way of studying the dynamics of the system

consists in analyzing the *vector field*, that is, the vector of temporal derivatives $(dx_a/dt, dx_b/dt)'$ associated with each point. This gives an indication of the direction of the trajectories passing through the point, as illustrated in (b). The analysis can be refined by tracing the *nullclines* in the phase plane, that is, the curves on which the temporal derivatives of x_a and x_b equal 0 (here, these curves are defined by $x_a = (\kappa_a/\gamma_a)\, h^-(x_b, \theta_b, m_b)$ and $x_b = (\kappa_b/\gamma_b)\, h^-(x_a, \theta_a, m_a)$). The points where the nullclines intersect are the *equilibrium points* of the system. If all trajectories in a neighborhood of the equilibrium point remain in that neighborhood, then the equilibrium point is *stable*. If, in addition, they converge towards the equilibrium point, the latter is *asymptotically stable*. So, by studying the vector field around the equilibrium point, one can determine its stability. In the case of the nonlinear model of the network in Fig. 3, there are three equilibrium points: two of these are asymptotically stable and one is unstable (Fig. 5). The result of the analysis summarized in this paragraph is often called the *phase portrait*.

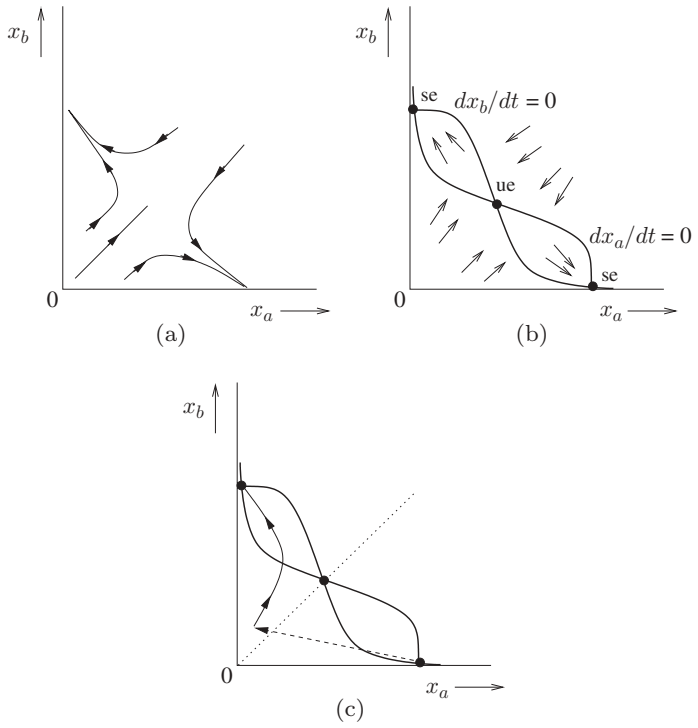


Fig. 5. Phase portrait of the nonlinear model of the mutual-inhibition network (Fig. 4). (**a**) Examples of trajectories. (**b**) Vector field and nullclines. The system has two asymptotically stable equilibrium points (se) and one unstable equilibrium point (ue). (**c**) Hysteresis effect, resulting from a transient perturbation of the system (*broken line with arrow*)

The above phase-plane analysis predicts that the mutual-inhibition network is *bistable*. That is, starting from certain initial conditions, the system will reach one of the two stable equilibria. From a practical point of view, the unstable equilibrium has no importance, because it is only attained for quite specific initial conditions. Moreover, a perturbation of the unstable equilibrium, even vanishingly small, will cause the system to converge towards one of the stable equilibria. The phase portrait also reveals that the system exhibits hysteresis. If one perturbs the system from one of its stable equilibria – for instance, by provoking a strong degradation of the protein present at a high concentration – the other equilibrium can be reached (Fig. 5(c)). From then onwards, even if the source of strong degradation has disappeared, the system will remain at the new equilibrium. In other words, the example suggests that a simple molecular mechanism may allow the system to switch from one functional mode to another. For this reason, mutual-inhibition networks, or more generally networks with positive feedback loops, have been assigned a central role in cellular differentiation [26].

It is important to remark that the above analysis is not just a theoretical exercise. In fact, the properties of the mutual inhibition network revealed by the analysis – bistability and hysteresis – have been experimentally tested by Gardner et al. [27]. The network of Fig. 3 has been reconstructed in *Escherichia coli* cells by cloning the genes on a plasmid. The genes have been chosen such that the activity of the corresponding proteins can be regulated by external signals. In addition, reporter genes have been added that allow the state of the cell to be measured. The resulting mutual-inhibition network functions independently from the rest of the cell, like a "genetic applet", in the words of the authors. Carefully-chosen experiments have shown that the system is bistable and can switch from one equilibrium to the other following chemical or heat induction.

The qualitative analysis of the dynamics of the mutual inhibition network, summarized in Fig. 5, is valid for a large range of parameter values. However, for certain parameter values, the behavior of the system changes, as can be verified in Fig. 6. By increasing the value of the parameter θ_b, the nullcline of x_a, defined by $x_a = (\kappa_a/\gamma_a)\, h^-(x_b, \theta_b, m_b)$, moves upwards. As a consequence, one of the stable equilibria and the unstable equilibrium approach and then annihilate each other. For values of θ_b close to, or above, κ_b/γ_b, the system loses its bistability and hysteresis properties. In the terminology of dynamical systems theory, a *bifurcation* has occurred [30].

Generally, for networks having more than two genes, an analysis in the phase plane is no longer possible. In certain cases, one can reduce the dimension of the system by simplifying the model, but most of the time, numerical techniques become necessary. *Numerical simulation* approximates the exact solution of the system of equations, by computing approximate values $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_m$ for $\boldsymbol{x}$ at consecutive time-points $t_0, \ldots, t_m$ (see [31] for an introduction). Many computer tools for numerical simulation have been developed, some specifically adapted to networks of molecular interactions (see

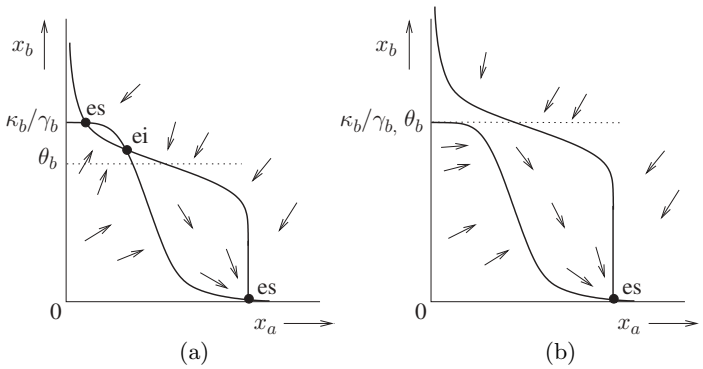


Fig. 6. Analysis of the bifurcation occurring when the value of the parameter θ_b is increased. The value in (**a**) is larger than the value in (**b**)

the references in [32]). These tools are at the heart of the analysis of nonlinear models of genetic regulatory networks. Unfortunately, their practical application is often difficult, due to the general absence of in vitro and in vivo measurements of the parameters of the model. These values are only available for a few systems whose functioning has already been well characterized experimentally.

Several solutions exist for dealing with the lack of quantitative data on the network components and their interactions. A first approach consists in using the increasing amounts of expression data, obtained by, for example, DNA microarrays or quantitative RT-PCR. Starting with measurements of the concentration variables $\boldsymbol{x}$ at several stages of the process under investigation in different experimental conditions, the parameter values can be estimated by means of *system identification* techniques [33]. A second approach consist in downplaying the importance of having precise value for the parameters. It is based on the hypothesis that it is the network structure rather than the parameter values that confers stability to the system. As a consequence, essential properties of the system should be robust to variations in parameter values, that is, for wide ranges of parameter values the model should reproduce the qualitative dynamics of the system [34, 35].

Many examples of the application of nonlinear differential equation models to prokaryote and eukaryote genetic regulatory networks can be found in the literature. To cite just a few, we mention the models of the infection of *E. coli* by the bacteriophages λ [8] and T7 [25], circadian rhythms in the fruit fly *Drosophila melanogaster* [7], and the establishment of segment polarity in the same organism [34].

3.2 Linear Models

Nonlinear ordinary differential equation models give an adequate description of important aspects of the dynamics of genetic regulatory networks.

Unfortunately, the nonlinear models become quite difficult to treat mathematically when passing from simple synthetic networks like the one in Fig. 3 to the complex networks controlling the functioning and development of real organisms. This raises the question whether the dynamics of genetic regulatory networks could not be equally well described by linear differential equation models, which possess more favorable mathematical properties.

A system of linear ordinary differential equations has the form (2), but the functions $\boldsymbol{f}$ are *linear*. That is, (2) can be rewritten as follows:

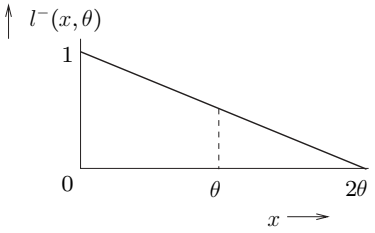
$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{A}\,\boldsymbol{x} + \boldsymbol{b},\ \ \boldsymbol{A} \in \mathbb{R}^{n\times n},\ \ \boldsymbol{b} \in \mathbb{R}^{n}\ . \tag{3}$$

Henceforward, we will make the hypothesis that the element of the matrix $\boldsymbol{A}$ and the vector $\boldsymbol{b}$ are constants. As a consequence, the system (3) has an analytical solution, given by linear systems theory [36].

How can we model a genetic regulatory network by means of linear ordinary differential equations? By way of example, the model of the mutual-inhibition network is shown in Fig. 7(a). It much resembles the nonlinear model presented in Sect. 3.1: as before, the time derivative is equal to the difference between a synthesis term and a degradation term. However, a linear function $l^- : D \times \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$, $D \subset \mathbb{R}_{\geq 0}$ is now used instead of the sigmoidal function h^-. As the latter function, l^- is monotonically decreasing, but it is characterized by a single parameter, θ, defining the slope. In addition, the domain of the variable x is restricted to $D = [0, 2\theta] \subset \mathbb{R}_{\geq 0}$, because $1 - x/(2\theta)$ becomes negative for $x > 2\theta$, thus violating the obvious constraint that the

$$\frac{dx_a}{dt} = \kappa_a\, l^-(x_b, \theta_b) - \gamma_a\, x_a$$
$$\frac{dx_b}{dt} = \kappa_b\, l^-(x_a, \theta_a) - \gamma_b\, x_b$$
$$l^-(x, \theta) = 1 - \frac{x}{2\theta}$$

(a)



(b)

$$\frac{d}{dt}\begin{bmatrix} x_a \\ x_b \end{bmatrix} = \begin{bmatrix} -\gamma_a & -\kappa_a/(2\theta_b) \\ -\kappa_b/(2\theta_a) & -\gamma_b \end{bmatrix} \begin{bmatrix} x_a \\ x_b \end{bmatrix} + \begin{bmatrix} \kappa_a \\ \kappa_b \end{bmatrix}$$

(c)

Fig. 7. (**a**) Linear ordinary differential equation model of the mutual-inhibition network (Fig. 3). The variables x_a and x_b correspond to the concentrations of the proteins A and B, respectively, the parameters κ_a and κ_b to the synthesis rates of the proteins, the parameters γ_a and γ_b to the degradation rates, and the parameters θ_a and θ_b to the strength of the interactions. All parameters are positive. (**b**) Graphical representation of the linear function $l^-(x, \theta)$. (**c**) Reformulation of the model in (**a**) in the matrix form of (3)

synthesis rate must be non-negative (Fig. 7(b)). It is easily verified that the model can be rewritten in the form (3) (Fig. 7(c)). Note that the model is only valid for $x_a \in [0, 2\theta_a]$ and $x_b \in [0, 2\theta_b]$, due to the definition of l^-.

As in the case of the nonlinear model, the qualitative dynamics of the network can be studied in the phase plane. Figure 8(a) shows some examples of trajectories in the phase plane. From a superficial comparison with Fig. 5(a), one would be inclined to conclude that the linear and nonlinear models make more or less identical predictions of the dynamics of the system. However, analysis of the nullclines – defined by $x_a = (\kappa_a/\gamma_a)\, l^-(x_b, \theta_b)$ and $x_b = (\kappa_b/\gamma_b)\, l^-(x_a, \theta_a)$ – shows that this is not the case (Fig. 8(b)). In fact, the linear system has only a single equilibrium point, corresponding to the unstable equilibrium point of the nonlinear system in Fig. 5(b). Almost all trajectories reach either one of the segments $x_a = 2\theta_a$ or $x_b = 2\theta_b$ after a while, and would continue towards $(-\infty, \infty)'$ or $(\infty, -\infty)'$, respectively, if

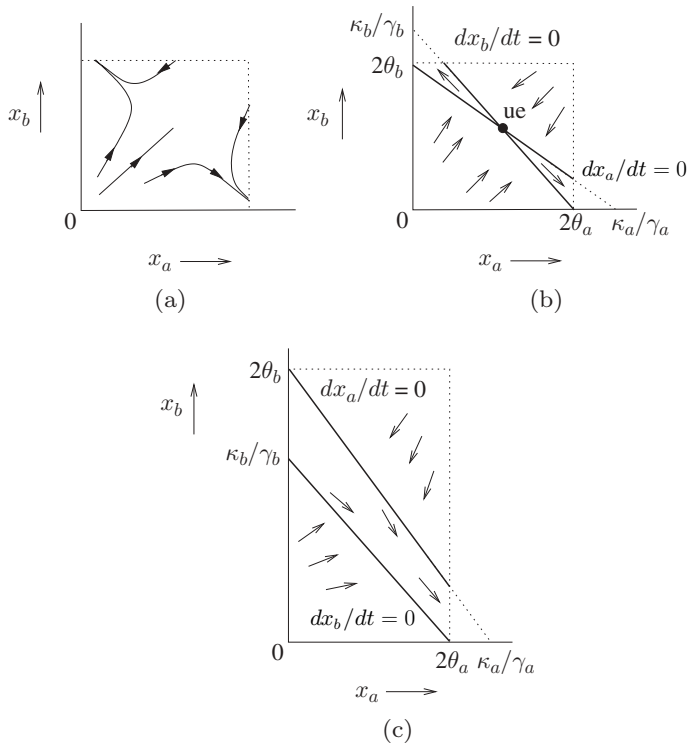


Fig. 8. Phase portrait of the linear model of the mutual-inhibition network (Fig. 7). (**a**) Examples of trajectories. (**b**) Vector field and nullclines. The system has a single unstable equilibrium point (ue). (**c**) Analysis of the bifurcation occurring when the value of the parameter θ_b is increased. The value of θ_b in (**c**) is larger than that in (**b**). The analysis is restricted to $[0, 2\theta_a] \times [0, 2\theta_b]$, the part of the phase space where the linear model is defined

the system were defined outside $[0, 2\theta_a] \times [0, 2\theta_b]$. Figure 8(c) shows that the equilibrium point disappears if one increases the value $2\theta_b$ above κ_b/γ_b, while keeping the other parameters at the same value. In that case, all trajectories reach the segment $x_a = 2\theta_a$.

The phase-plane analysis summarized in Fig. 8 teaches us that, when modeled by a system of linear differential equations, the mutual-inhibition network no longer exhibits bistability or hysteresis. The predictions of the model therefore contradict what is experimentally observed by Gardner and colleagues [38]. In fact, the example shows that the nonlinear character of the inhibition of gene expression by regulatory proteins, expressed by means of the function h^-, is crucial for the global dynamics of the network. The approximation of h^- by l^- is unable to preserve essential properties of the dynamics. On the other hand, the analysis of the two-gene network suggests that linear models could contribute to the analysis of the local dynamics of the system. For example, even though they do not converge towards a stable equilibrium point, the trajectories in Fig. 8(a) resemble those predicted by the nonlinear model in the neighborhood of the unstable equilibrium (Fig. 5(a)).

This property of linear models can be exploited when trying to reconstruct the connectivity of a genetic regulatory network from experimental data. Suppose one had a time-series of measurements of the concentration variables, obtained by DNA microarrays or quantitative RT-PCR. This series of measurements can be represented in the form of a matrix $\hat{\boldsymbol{X}}$, where $\hat{\boldsymbol{X}} \in \mathbb{R}^{n \times m}$. Every element $\hat{x}_{ij}$ of this matrix represents a measurement, more specifically the measurement of the variable x_i at time-point j. Instead of a time-series of measurements, the columns of the matrix $\hat{\boldsymbol{X}}$ could also represent measurements realized under various experimental conditions, for instance, the steady state reached by a mutant of the organism after a physiological perturbation.

System identification techniques [33] allow the values of the elements of the matrix $\boldsymbol{A}$ and the vector $\boldsymbol{b}$ to be estimated from measurements $\hat{\boldsymbol{X}}$. These estimations make it possible to infer the interaction structure of a network, as can be easily understood by considering the matrix $\boldsymbol{A}$ in the case of the mutual-inhibition network (Fig. 7(c), see also [23, 37]). In fact, the negative sign of the off-diagonal elements a_{ab} and a_{ba} corresponds to the inhibition of gene a by protein B, and the inhibition of gene b by protein A, respectively. If B activated a and A activated b, these elements would have been positive (which can be simply verified by replacing $l^-(x, \theta)$ by $l^+(x, \theta) = 1 - l^-(x, \theta)$ in the model of Fig. 7(a)). More generally, it follows that the estimation of the values of $\boldsymbol{A}$ and $\boldsymbol{b}$ from expression data provides us with information about the regulatory structure of the system, that is, on the existence of interactions between the genes and the nature of these interactions (activation, inhibition).

From a technical point of view, the use of linear ordinary differential equation models simplifies the approach of reconstructing genetic regulatory networks from gene expression data. In comparison with nonlinear models, linear models have a restricted number of parameters, as can be seen by comparing the models in the Figs. 4(a) and 7(a). In addition, powerful techniques for

parameter estimation exist for linear models. Taken together, this makes linear models more adapted to the quantitative and qualitative limitations of the experimental data available today. In fact, the expression data obtained by DNA microarrays are often noisy and the number of measurements m much smaller than the number of variables n. Therefore, most studies on network reconstruction to date have used linear or pseudo-linear models. Examples of such studies are the reconstruction of the SOS regulon in *E. coli* [38, 39], part of the network underlying rat central nervous system development [12], and a modular view of the global regulatory network in *S. cerevisiae* [11].

3.3 Piecewise-Linear Models

The linear differential equation models are easier to analyze than the nonlinear models, as we have seen in Sect. 3.2. However, this mathematical simplicity comes at the price of a reduced ability to take into account essential properties of the dynamics of the system. Are there models that are easy to treat mathematically and nevertheless capable of adequately representing the dynamics of the system? In this section, we will study a class of models that answer both requirements: piecewise-linear differential equations.

The general form of the models is given by (2), with the additional constraint that the functions $\boldsymbol{f}$ are *piecewise linear*. That is, the phase space $\mathbb{R}^n_{\geq 0}$ is divided into regions $\Delta^{(j)}$, $j \in [1, \ldots, p]$, in each of which the network is described by a system of linear differential equations. While being *globally* nonlinear, a piecewise-linear differential equation model is *locally* linear.

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{A}^{(j)}\,\boldsymbol{x} + \boldsymbol{b}^{(j)} \tag{4}$$

$$\boldsymbol{A}^{(j)} \in \mathbb{R}^{n\times n},\ \boldsymbol{b}^{(j)} \in \mathbb{R}^n,\ \boldsymbol{x} \in \Delta^{(j)} \subseteq \mathbb{R}^n_{\geq 0},\ j \in [1, \ldots, p]\,.$$

As for the linear models, we will assume that the elements of $\boldsymbol{A}^{(j)}$ and $\boldsymbol{b}^{(j)}$ are constants. This implies that in each region $\Delta^{(j)}$, (4) can be solved analytically.

Piecewise-linear differential equations have been used to model genetic regulatory networks since the early seventies [40, 41, 42, 43]. In order to illustrate their application, we will again consider the example of the two-gene network. The piecewise-linear model, presented Fig. 9, is obtained from the nonlinear model by replacing the sigmoidal function h^- by another approximation, the step function $s^- : D \times \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$, $D \subset \mathbb{R}_{\geq 0}$. For concentrations x below the threshold θ, $s^-(x, \theta)$ equals 1, whereas for concentrations x above θ, the function evaluates to 0. For $x = \theta$, it is not defined. As one can verify in Fig. 9(c), the model can be rewritten in the form (4). The segments $x_a = \theta_a$ and $x_b = \theta_b$ divide the phase space into four regions, $\Delta^{(1)}, \ldots, \Delta^{(4)}$. In each region, after evaluation of the step functions, the model reduces to a system of two linear differential equations.

In the cases that interest us, the reduced system of differential equations associated with a region $\Delta^{(j)}$ is not only linear, but also *uncoupled*. That

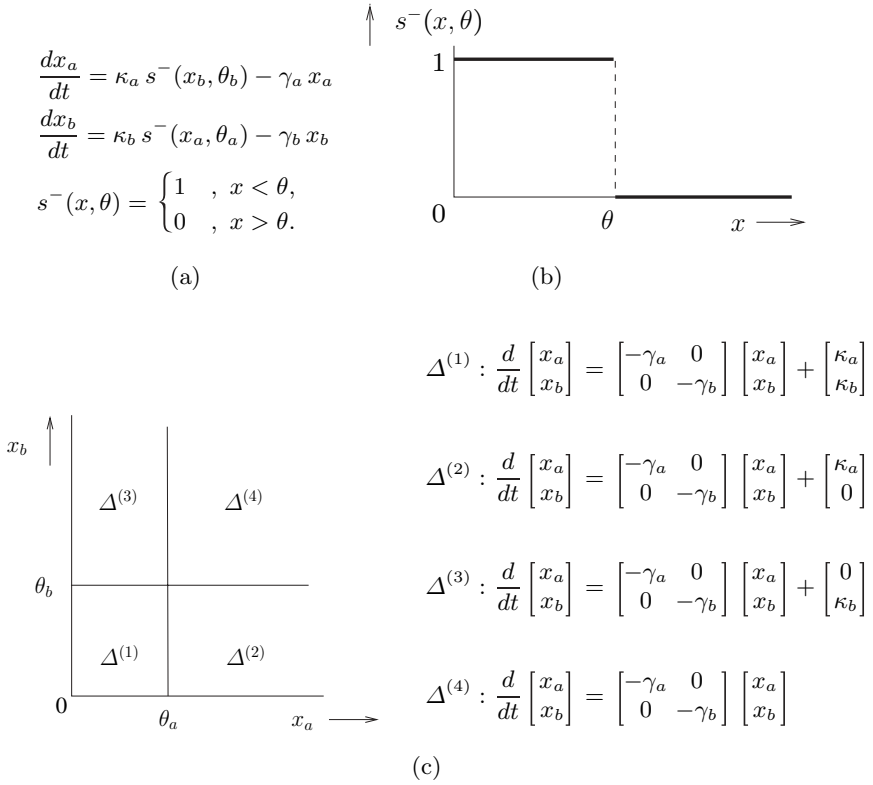


Fig. 9. (**a**) Piecewise-linear differential equation model of the mutual-inhibition network (Fig. 3). The variables x_a and x_b represent the concentration of proteins A and B, respectively, the parameters κ_a and κ_b the synthesis rates, the parameters γ_a and γ_b the degradation rates, and the parameters θ_a and θ_b threshold concentrations for A and B, respectively. All parameters are positive. (**b**) Graphical representation of the step function $s^-(x,\theta)$. (**c**) Piecewise-linear structure of the model in (**a**), corresponding to the division of the phase space into four regions $(\Delta^{(1)}, \ldots, \Delta^{(4)})$ by $x_a = \theta_a$ and $x_b = \theta_b$

is, $\boldsymbol{A}^{(j)}$ is a diagonal matrix and the temporal derivative of the variable x_i does not depend on variables other than x_i. Such a system has a very simple analytical solution. In fact, one can show that, in the region $\Delta^{(j)}$, all solutions locally converge towards the point $\phi(\Delta^{(j)}) = (b_1^{(j)}/a_{11}^{(j)}, \ldots, b_n^{(j)}/a_{nn}^{(j)})'$ [40]. For instance, in $\Delta^{(1)}$, the solutions converge towards $\phi(\Delta^{(1)}) = (\kappa_a/\gamma_a, \kappa_b/\gamma_b)'$, while in $\Delta^{(2)}$, they converge towards $\phi(\Delta^{(2)}) = (\kappa_a/\gamma_a, 0)'$ (Fig. 10(b)). If $\phi(\Delta^{(j)}) \in \Delta^{(j)}$, then $\phi(\Delta^{(j)})$ is an equilibrium point of the system, which is for instance the case for $\Delta^{(2)}$ and $\Delta^{(3)}$.

The piecewise-linear model does not specify how the system behaves on the segments $x_a = \theta_a$ and $x_b = \theta_b$, where one or more step functions, and

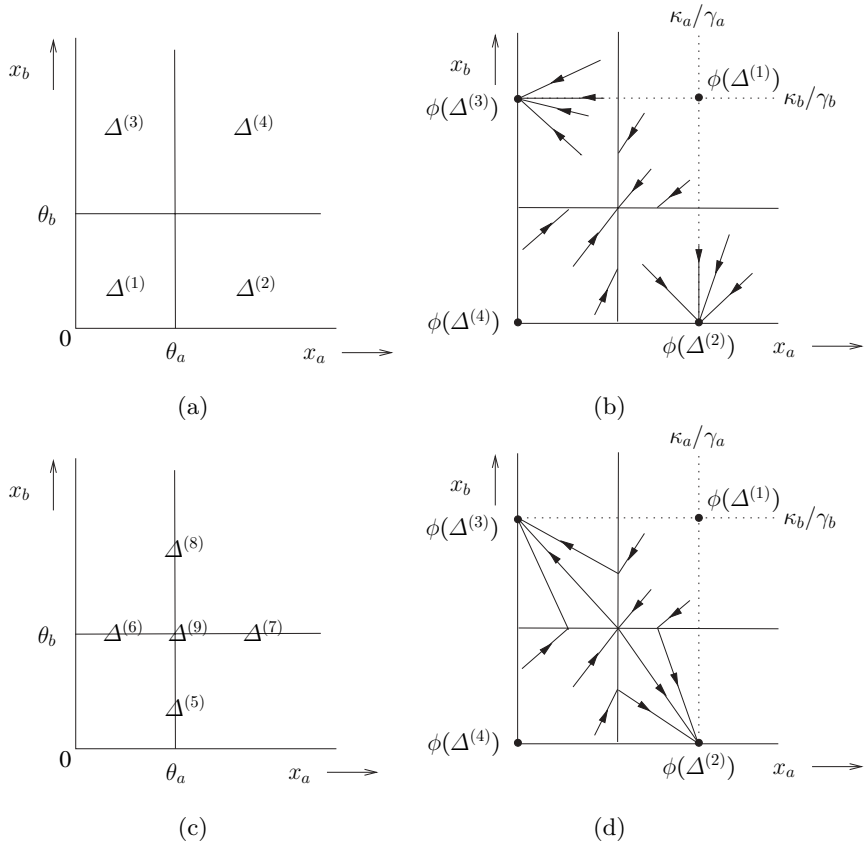


Fig. 10. Local phase portraits of the piecewise-linear model of the mutual-inhibition network (Fig. 9). The figure shows (**a**) the regions $\Delta^{(1)}, \ldots \Delta^{(4)}$, (**b**) examples of trajectories in these regions, (**c**) the regions $\Delta^{(5)}, \ldots, \Delta^{(9)}$ located on the segments $x_a = \theta_a$ or $x_b = \theta_b$, and (**d**) examples of trajectories arriving at or departing from these regions. The trajectories are straight lines, because in the simulations we set $\gamma_a = \gamma_b$

hence the corresponding differential equations, are not defined. In order to treat this problem, Gouzé and Sari [44] have proposed an approach which consists of extending the differential equation model (4) to a differential inclusion model, following ideas developed in control theory. This solution exploits mathematical concepts that are outside the scope of this chapter, but for our purposes, it is sufficient to know that the approach is elegant from a theoretical point of view and easy to use in practice. In the example, it allows the local analysis of the dynamics of the network to be extended to regions of the phase space located on the segments $x_a = \theta_a$ and $x_b = \theta_b$, that is, $\Delta^{(5)}, \ldots, \Delta^{(9)}$ (Fig. 10(c)). The results of the analysis are intuitive: the solutions of the system instantaneously traverse $\Delta^{(5)}, \ldots, \Delta^{(8)}$, whereas solutions

reaching $\Delta^{(9)}$ can remain indefinitely in this region or leave it after a certain time (Fig. 10(d)).

The local analyses of the dynamics of the system in the different regions of the phase space can be combined into a global analysis, as illustrated in Fig. 11. The predictions of the piecewise-linear model are qualitatively equivalent to those obtained by the nonlinear model. The network has three equilibrium points, of which two are stable and one is unstable (Fig. 11(a)). Part (c) of the figure shows that a transient perturbation may cause the

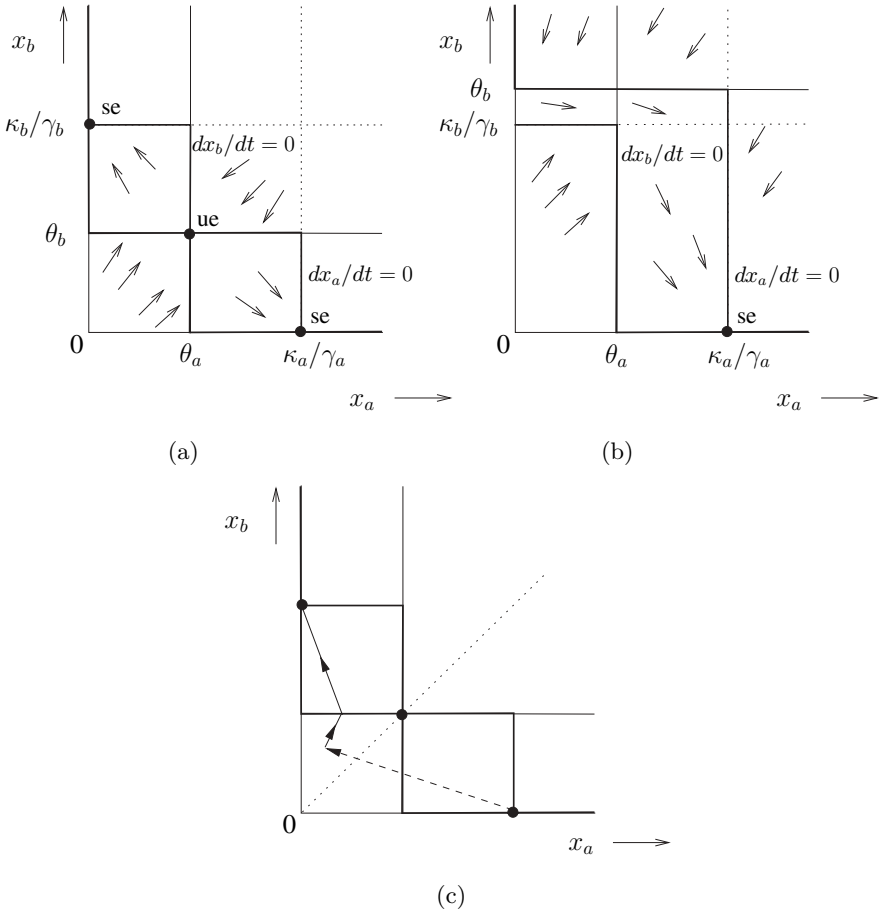


Fig. 11. Global phase portrait of the piecewise-linear model of the mutual-inhibition network (Fig. 9). (**a**) Vector field and nullclines. The system has two stable equilibrium points (se) and one unstable equilibrium point (ue). (**b**) Analysis of the bifurcation produced when the value of the parameter θ_b is increased. The value of θ_b in (**b**) is larger than that in (**a**). (**c**) Hysteresis phenomenon, following a transient perturbation of the system (*broken line with arrow*)

system to switch from one stable equilibrium to the other. As for the non-linear model, an increase of the value of the parameter θ_b, without changing the value of the other parameters, can bring about a bifurcation: one of the two stable equilibria and the unstable equilibrium disappear (Fig. 11(b)). In summary, the example shows that, while facilitating the mathematical analysis, the piecewise-linear models allow us to preserve essential properties of the mutual-inhibition network. There are good reasons to believe that this is also true for other, more complex networks, but this has not been formally proven yet.

The analysis of the piecewise-linear model of the two-gene network suggests a discrete, more compact representation of the dynamics of the system [45]. In fact, every region of the phase space can be seen as a *qualitative state*, in which the system behaves in a qualitatively homogeneous way. For instance, in the region $\Delta^{(1)}$, all trajectories converge towards the point $\phi(\Delta^{(1)}) = (\kappa_a/\gamma_a, \kappa_b/\gamma_b)'$, whereas in $\Delta^{(2)}$, they converge towards $\phi(\Delta^{(2)}) = (0, \kappa_b/\gamma_b)'$. Two qualitative states can be connected by a *transition*, if there exists a solution starting in the region corresponding to the first state that reaches the region corresponding to the second state, without passing through a third region. This is the case for the solutions in $\Delta^{(1)}$ which, while converging towards $\phi(\Delta^{(1)})$, reach $\Delta^{(5)}$, $\Delta^{(6)}$ or $\Delta^{(9)}$. The set of qualitative states and transitions between these states defines a *state transition graph*.

The state transition graph obtained for the model of the mutual-inhibition network is shown in Fig. 12(a). The graph is composed of nine qualitative states, associated to the regions of the phase space (Fig. 10), and the transitions between these states. Three of the nine states are *qualitative equilibrium states*, that is, states corresponding to a region containing an equilibrium point. The graph summarizes the dynamics of the network in a qualitative manner. For instance, it provides information on the reachability of an equilibrium point from a given region. If the equilibrium point is reachable, there must exist a path in the graph going from the qualitative state corresponding to the initial region to the qualitative equilibrium state corresponding to the region in which the equilibrium point is contained.

Generally speaking, the state transition graph associated to a piecewise-linear model will vary with the parameter values. However, following de Jong and colleagues [45], one can define a class of models determined by inequality constraints on the parameters. Under certain, not too restrictive conditions, each model in that class produces the same state transition graph. This can be illustrated by means of the example of the two-gene network, by considering the transitions from QS^1, the qualitative state corresponding to the region $\Delta^{(1)}$. The transitions from QS^1 to the states QS^5, QS^6 and QS^9 do not depend on the exact values of the parameters, as long as $\kappa_a/\gamma_a > \theta_a$ and $\kappa_b/\gamma_b > \theta_b$. In fact, under these conditions, $\phi(\Delta^{(1)}) \in \Delta^{(4)}$ and the trajectories in $\Delta^{(1)}$ all reach QS^5, QS^6 or QS^9 after a certain time. A *qualitative simulation* method has been proposed, which symbolically computes the state transition graph

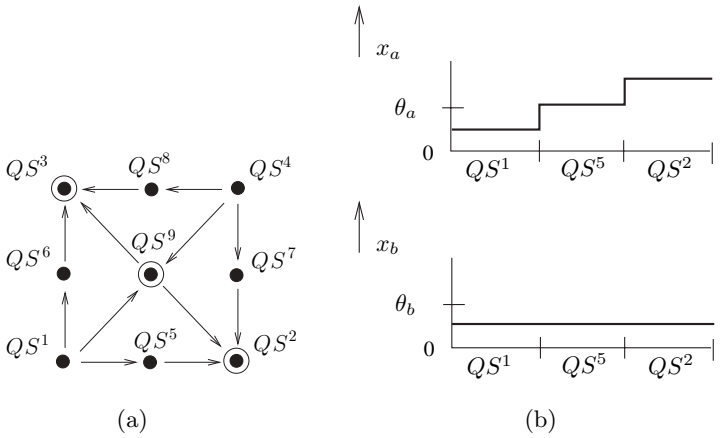


Fig. 12. (**a**) State transition graph produced from the piecewise-linear model of the mutual-inhibition network (Fig. 9). The qualitative equilibrium states are *circled* [45]. (**b**) Detailed description of the sequence of qualitative states $\langle QS^1, QS^5, QS^2 \rangle$

for a piecewise-linear differential equation model, supplemented by inequality constraints on the parameters [45]. This method has been implemented in a computer tool called Genetic Network Analyzer (GNA) [47].

The interest of qualitative simulation derives from the fact that it is adapted to the lack of quantitative information on genetic regulatory networks, a problem already referred to in previous sections. Instead of numerical values, the method uses inequality constraints that can usually be specified by means of the qualitative information available in the experimental literature. On the formal level, the qualitative simulation method is related to a method developed by Thomas and colleagues, which is based on asynchronous logical models [26, 48]. The two methods have demonstrated their usefulness in the study of a certain number of prokaryotic and eukaryotic networks, whose analysis is rendered difficult by the almost complete absence of numerical parameter values ([46, 49, 50, 51, 52]; see [53] for a review and [3, 4] for related ideas).

4 Coupled Map Networks

In this section, we consider another class of models, namely discrete time systems with continuous variable. As shown during this school discrete time models are widely employed to model the dynamics of spatially extended systems, notably in Physics. Excepted for logical models, discrete time models of genetic regulation networks however are somehow usual.

As argued below, discrete-time models provide a simple framework where to analyse the consequences of interactions delays on the dynamics. They also largely benefit from tools and techniques in the theory of dynamical systems.

To be specific, the models considered in this section consist of iterations of network mappings, the so-called *Coupled Map Networks* (CMN). For the sake of simplicity, we only consider piecewise affine and contracting CMN. The dynamics, i.e. the orbits $\{x^t\}_{t\in\mathbb{Z}}$, is given by the following induction (piecewise-affine and contracting) [57, 58]

$$x_i^{t+1} = ax_i^t + \sum_{j\in I(i)} K_{ij} H(s_{ij}(x_j^t - T_{ij})) \; . \tag{5}$$

In this expression, the subscript i runs over $\{1,\ldots,N\}$ for a network composed of N nodes. Each variable x_i^t belongs to the interval $[0,1]$ and represents a (density of) concentration at time t. The parameter a belongs to $[0,1)$ and represents an extinction rate (in absence of any other interaction). The symbol H denotes the Heaviside function:

$$H(x) = \begin{cases} 1 & \text{if} \quad x \geq 0 \\ 0 & \text{if} \quad x < 0 \end{cases} \; .$$

In relation (5), the set $I(i)$ is the set of genes which have an action over i. In other terms, instead of being defined by an adjency matrix, the interaction graph associated with the network is entirely characterised by the collection of sets $\{I(i)\}_{i=1}^N$.

The action from j to i is either an activation when the number s_{ij}, called the sign, is equal to $+1$, or an inhibition when $s_{ij} = -1$. The corresponding weights K_{ij} are assumed to be positive according to the assumption of cumulative interactions. For the sake of simplicity and without loss of generality [57] the weights can always be assumed to be normalized as follows

$$\sum_{j\in I(i)} K_{ij} = 1 - a.$$

Finally, the threshold parameters T_{ij} belong to the open interval $(0,1)$. (These parameters were denoted by the symbol θ in the previous section.) For the sake of coherence with the references [54, 56, 57], we keep the notation θ for the symbols associated with the atoms of the partition, see below.

The choice of a piecewise affine model not only simplifies the analysis but also allows to globally describe the dynamics both in phase space (here the set $\mathbb{R}^N$) and parameter space. However, by using continuation arguments some results on the existence and stability of orbits can be shown to extend to smooth perturbations.

Along the same line, that the synthesis term in (5) is a linear combination of basic interactions $H(s_{ij}(x_j - T_{ij}))$ is another simplifying assumption. According to the genetic regulatory network under consideration, other synthesis terms, such as combinations of products of basic interaction or more general nonlinear interactions may be more relevant. Once these terms remain piecewise constant, the global mapping remains piecewise affine and the

principle of analysis still apply. Naturally, since the phase space partition into domains with affine dynamics gets more elaborated, the global description of the dynamics and its changes with parameters is getting more involved.

According to (5), each iteration of a given gene concentration x_i can be viewed as the action of a one-dimensional affine map chosen among $2^{I(i)}$ maps (see Fig. 13). Each affine map corresponds to a choice of $\sum_{j\in I(i)} K_{ij}\theta_{ij}$ where $\theta_{ij} \in \{0, 1\}$ are called *symbols*.

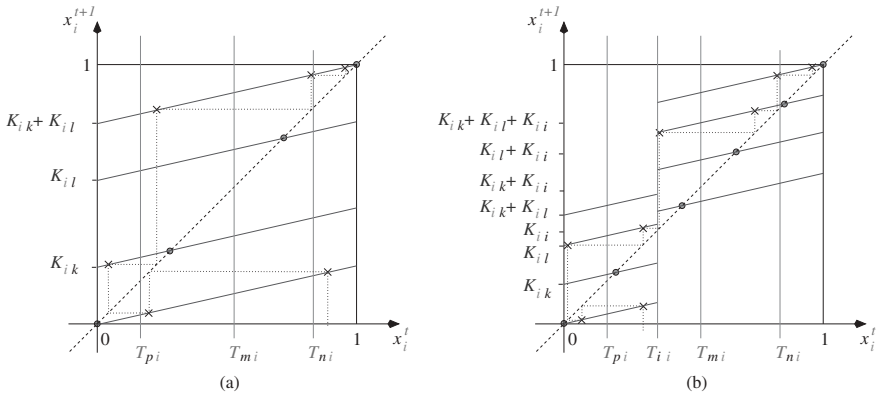


Fig. 13. The one-dimensional affine maps acting on a given gene concentration x_i in the case where (**a**) $I(i) = \{k, l\}$, (**b**) $I(i) = \{i, k, l\}$. In complement, we have also assumed that $i \in I(p)$, $i \in I(m)$ and $i \in I(n)$

There are two cases depending on the fact that the given gene i has self-interaction (i.e. when $i \in I(i)$) or not (i.e. when $i \notin I(i)$). In the second case, the choice of the value of the symbol does not involve the gene itself and each of the previous one-dimensional map is defined on the whole interval $[0, 1]$.

In the first case however choosing a value for the symbol θ_{ii} implicitly assumes that $x_i < T_{ii}$ or $x_i \geq T_{ii}$[1]. Therefore all one-dimensional maps are only defined on some interval (the interval for which $\theta_{ii} = 0$ or the one for which $\theta_{ii} = 1$). Nonetheless, the maps can be grouped pairwise in order to define discontinuous maps defined on the whole $[0, 1]$, each map corresponding to a choice of $\{\theta_{ij}\}$ for $j \neq i$ and letting θ_{ii} be arbitrary (see Fig. 13 (b)).

In any case, each affine one-dimensional map has a fixed point defined by

$$\frac{1}{1-a} \sum_{j\in I(i)} K_{ij}\theta_{ij} \; . \tag{6}$$

By normalisation, this point belongs to $[0, 1]$ (see Fig. 13).

As the simplest feature of a dynamical system, one may first be interested in fixed points of the CMN. The previous arguments on one-dimensional maps

[1] $x_i \leq T_{ii}$ or $x_i > T_{ii}$ if $s_{ii} = -1$

at each gene show that the n-uple $(x_1, \cdots, x_n)$ is a fixed point of the system (5) only if each of its component x_i is a fixed of a corresponding one-dimensional map, i.e. if we have $x_i = \frac{1}{1-a} \sum_{j \in I(i)} K_{ij} \theta_{ij}$.

Once all these components have been chosen, actually once the value of the symbols $\{\theta_{ij}\}$ have been chosen, one has to ensure that $(x_1, \cdots, x_n)$ is invariant under the action of the CMN. A sufficient condition for this is

$$\theta_{ij} = H(s_{ij}(x_j - T_{ij})) \quad \forall i = 1, \ldots, N, \ \forall j \in I(i)$$

i.e. the result of computing $H(s_{ij}(x_j - T_{ij}))$ with the fixed point component gives the prescribed symbol value θ_{ij}. Geometrically, this condition imposes that the fixed point resulting from the choice of symbol values $\{\theta_{ij}\}$ belongs to the corresponding domains.

If this condition is satisfied, then the prescribed fixed point exists. Otherwise it does not exist. Naturally, the existence of a fixed point depends on the network structure and on the parameters a, K_{ij} and T_{ij}.

The reasoning which consists firstly in computing a priori the orbit components using symbols and secondly in checking that the computed orbit actually belongs to the prescribed region is crucial in piecewise affine dynamical systems [54]. In particular, it is by applying this method that most results on CMN have been obtained.

In order to describe other orbits than fixed points in the CMN, for simplicity we start by considering those orbits in the attractor. By a technical argument, it turns out that the orbits in the attractor exactly correspond to orbits with infinite (and bounded) past, the so-called *global orbits* [57]. Precisely, a global orbit is a bounded sequence $\{x^t\}_{t \in \mathbb{Z}}$ which satisfies the relation (5) for all $t \in \mathbb{Z}$. An orbit belongs to the attractor of the CMN iff it is global.

Given an arbitrary orbit $\{x^t\}$ of the CMN (not necessarily a global one) a symbolic sequence $\{\theta_{ij}^t\}$, namely its *code*, can be computed. For each t, each $i \in \{1, \ldots, N\}$ and each $j \in I(i)$, we compute the symbol associated with the interaction from j to i at time t, i.e.

$$\theta_{ij}^t = H(s_{ij}(x_j^t - T_{ij})) \quad i \in \{1, \ldots, N\}, \ j \in I(i) \tag{7}$$

On the other hand, the geometric induction in (5) can be solved to obtain the expression of an arbitrary orbit. In particular, for a global orbit, this expression only depends on the code and we have

$$x_i^t = \sum_{j \in I(i)} K_{ij} \sum_{q=0}^{\infty} a^q \theta_{ij}^{t-q-1} \quad i \in \{1, \ldots, N\}, \ t \in \mathbb{Z} \, . \tag{8}$$

Notice that in the case where the code is constant $\theta_{ij}^t = \theta_{ij}$ for all t, this expression reduces to the expression (6) of fixed points. More generally, relation (8) gives the expression of any periodic orbit by choosing a periodic code $\theta_{ij}^{t+T} = \theta_{ij}^t$, the expression of any quasi-periodic orbit when a quasi-periodic code is chosen and so on.

Since $a \in [0,1)$, any fixed point is locally stable if none of its component intersects the (corresponding) interaction thresholds (if $x_j \neq T_{ij}$ for all i,j). However (and contrary to as in logical models) when $a \neq 0$, fixed points cannot (generically) be reached in a finite number of iterations. To a broader extent, orbits do not reach stable periodic orbits in a finite number of iterations and it may happen that the trajectory depends entirely on its past history.

In short terms, we have shown that an orbit belongs to the attractor of the CMN iff 1/ its components are given by (8) and 2/ its code is given by (7). This is the fundamental statement employed to describe the symbolic dynamics; namely guess a symbolic sequence, compute the orbit and check that the symbolic sequence is actually the code of the computed orbit.

Alternatively, the previous statement can be phrased in terms of symbolic sequences. A symbolic sequence $\{\theta_{ij}\}_{t\in\mathbb{Z}}$ is the code of an orbit in the attractor iff it is given by (7) and the components are given by (8). At once, a symbolic sequence $\{\theta_{ij}\}_{t\in\mathbb{Z}}$ is the code of an orbit in the attractor iff it satisfies the *admissibility condition*

$$\theta_{ij}^t = H\left(s_{ij}\left(\sum_{l\in I(i)}\sum_{q=0}^{\infty} a^q \theta_{jl}^{t-q-1} - T_{ij}\right)\right) \quad i \in \{1,\dots,N\},\ j \in I(i),\ t \in \mathbb{Z}\,.$$

The admissibility condition is nothing but the condition that the components of the global orbit corresponding to a symbolic sequence belong to the region of phase space specified by the symbols. It is the analogous for an arbitrary global orbit of the condition above for the existence of a fixed point with a prescribed code, namely

$$\theta_{ij} = H\left(s_{ij}\left(\frac{1}{1-a}\sum_{l\in I(i)} K_{jl}\theta_{jl} - T_{ij}\right)\right) \quad \forall i = 1,\dots,N,\ \forall j \in I(i)$$

It is important to notice that computing x_i^t uses the symbols $\{\theta_{ij}^t\}_{j,t}$ whereas checking that x_i^t belongs to a suitable atom uses the symbols $\{\theta_{ji}^t\}_{j,t}$. Moreover, while the variables x_i are associated with genes, the symbols θ_{ij} are associated with interactions between genes.

Describing the attractor and its changes with parameters amounts to describe the set of admissible sequences and its changes with parameters. Technically, we have shown that the CMN on its attractor is conjugated with the action of the left shift on admissible symbolic sequences.

We now show that the orbits generally persist under small changes of parameters. The required condition is that they differ from any threshold parameter. Actually, they must be bounded away from these thresholds. To be precise, suppose that an admissible symbolic sequence $\{\theta_{ij}^t\}$ is given for some values of the parameters a, K_{ij} and T_{ij}. Suppose also that the distance between the components and the corresponding thresholds is uniformly bounded from 0, i.e.

$$\inf_{i\in\{1,\dots,n\},\ j\in I(i),\ t\in\mathbb{Z}} |x_j^t - T_{ij}| = \delta > 0.$$

(The sequence is said to be *strictly admissible*.) Then, the thresholds parameters can be slightly modified without affecting the orbit. That is to say, there exist[2] $T'_{ij} \sim T_{ij}$ for all i, j such that

$$\theta_{ij}^t = H(s_{ij}(x_j^t - T'_{ij})) \quad i \in \{1, \dots, N\},\ j \in I(i), t \in \mathbb{Z}$$

with θ_{ij}^t and x_i^t kept unchanged. Consequently, the same code remains admissible for the CMN with parameters a, K_{ij} and T'_{ij} with the orbit unchanged.

We have shown that the orbits persist under small changes in threshold. By using continuity of the components (8) with the parameters a and K_{ij}, one can shown that the same code remains admissible for the CMN with parameters a', K'_{ij} and T'_{ij} sufficiently close to the original ones. (If $a' \neq a$ and/or $K'_{ij} \neq K_{ij}$ then the orbit component will be slightly modified.)

On one hand, this shows that each global orbit is expected to exist on open sets of parameters (on product of intervals in threshold space when a and K_{ij} are kept constant). On the other hand, it shows that a qualitative change in the attractor (a bifurcation) only happens when some global orbit component passes an interaction threshold.

Numerical simulations of networks with a large number of genes reveal that the asymptotic dynamics very often consists of clusters of genes with oscillating behaviour immersed in sea of genes in a stationary state [55]. The size and distribution of clusters usually depends on the initial condition. Note that from relation (5), a gene concentration x_i^t oscillates only if an incoming symbol θ_{ij}^t oscillates with $j \in I(i)$. That is to say, x_i^t oscillates only if the concentration x_j^t oscillates for some gene j acting on i. By a recursive argument, there must be a closed path in the network graph on which all genes have oscillating concentration genes must belong to a circuit of genes with oscillating concentrations. A closed oriented path passing once through its nodes is called a *circuit* (a feedback loop in the previous section). A network graph which consists in a unique circuit (i.e. up to a relabeling of genes, we have $I(i) = i - 1 \bmod N$) is said to be an isolated circuit.

A closer examination [59] shows that the presence of oscillations in a circuit depends on the product of signs s_{ij} along the circuit. In positive circuit (i.e. when the product equals 1) the oscillations occur for some initial conditions and some parameters. In negative circuits, oscillations occur for all initial conditions and all parameters. (A complementary result shows that, in a isolated circuit, if one of the thresholds were chosen outside the interval $[0, 1]$, then any orbit would converge toward a fixed point. In particular, no oscillation can occur in this case.)

As building bricks of larger networks, isolated circuits are interesting in their own. By using symmetries, one shows that two circuits with N genes and identical product of signs are conjugated[3] by a simple change of variables and

[2] Actually any T'_{ij} such that $|T'_{ij} - T_{ij}| < \delta$ will do.

[3] Precisely, the orbits outside discontinuities are conjugated.

parameters. Therefore, in practice, one only has to investigate the dynamics of a representative of positive (resp. negative) circuit for any $N \in \mathbb{N}$.

As in Sects. 3.1 and 3.3, we now investigate the dynamics of the two-gene network with mutual inhibition. (A formal way to define this network in the present formalism is by letting $n = 2$, $I(1) = 2$, $I(2) = 1$ and $s_{12} = s_{21} = -1$.) Using also the normalisation $K_{12} = K_{21} = 1 - a$, the CMN becomes in this case

$$\begin{cases} x_1^{t+1} = ax_1^t + (1-a)\theta_{12}^t & \text{where } \theta_{12}^t = H(T_{12} - x_2^t) \\ x_2^{t+1} = ax_2^t + (1-a)\theta_{21}^t & \text{where } \theta_{21}^t = H(T_{21} - x_1^t) \end{cases} \tag{9}$$

where again H is the Heaviside function.

As in Sect. 3.3, the thresholds T_{21} and T_{12} define a partition of phase space into 4 atoms (see Fig. 14). (According to the definition of H, the CMN is well-defined on discontinuities lines which are included in the 4 atoms.) Naturally, the direction of motions in each atom are the same as for the differential equations. We shall see that a large part of the dynamics is also the same.

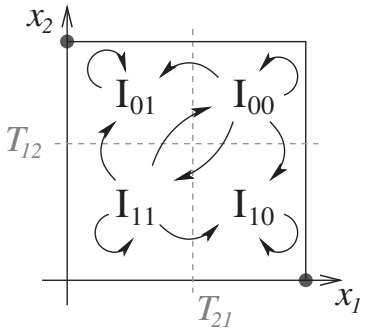


Fig. 14. Partition of the phase space for the mutual inhibition of two genes. The domains (atoms) labels $I_{\theta_{21}\theta_{12}}$ follow coding and the *arrow* show the possible transition between atoms

Firstly and just as for the differential equation, the attractor is contained in a square, the square $[0,1]^2$ in our case according to weight normalisation. It means that this square is invariant and that the orbit issued from any initial condition $(x_1^0, x_2^0) \in \mathbb{R}^2$ asymptotically approaches $[0,1]^2$. Actually, this property is not limited to this network but extends to arbitrary CMN [57].

Therefore, one may restrict attention to initial conditions and to orbits lying in the square $[0,1]^2$.

Secondly and as mentioned above, a displacement of any threshold outside $[0,1]$ results in a (unique) globally attracting fixed point and the system is monostable (see also Figs. 6 and 11).

Keeping the initial assumption $T_{12}, T_{21} \in (0,1)$, Fig. 14 shows that I_{01} and I_{10} are invariant. Due to contraction (i.e. $|a| < 1$), they both contain a unique fixed point $(0,1)$, and $(1,0)$ respectively, which attracts all initial condition in I_{01} (resp. I_{10}).

Moreover, points in I_{11} (resp. I_{00}) can be mapped in any atom I_{01}, I_{10}, I_{00} or I_{11}. In particular, the set of points in I_{11} which are mapped in I_{00} (and vice-versa) is a rectangle. This property is in contrast with differential equations where only special initial conditions (those on a specific line) can reach simultaneously both discontinuities, see Fig. 10.

As in differential equations, every orbit in I_{11} must leave this atom after a finite transient. If the orbit enters I_{01} or I_{10}, then it converges to the corresponding fixed point. If it enters I_{00}, then it may stay there for a while but must leave this atom after a finite transient. Again, either it enters I_{01} or I_{10} and it converges to a fixed point, or it enters I_{11}.

Hence the dynamics of orbits eventually entering one of the atoms I_{01} or I_{10} is just as for differential equations. The orbits converge to one of the fixed point $(0,1)$ or $(1,0)$. However and depending on parameters, there are orbits which oscillate forever between I_{00} and I_{11}.

If such orbits exist, then their code has identical symbols at each time, i.e. $\theta_{12}^t = \theta_{21}^t$ for all t. Due to contraction of the CMN, the orbit approach (or coincide) with some global orbits with the same code property. According to (8), every global orbit with $\theta_{12}^t = \theta_{21}^t$ for all t lie on the diagonal, i.e. we have $x_1^t = x_2^t := x^t$ for all t and where x^t necessarily satisfies the following induction

$$x^{t+1} = ax^t + (1-a)H(T_{12} - x^t) = ax^t + (1-a)H(T_{21} - x^t)$$

By applying results for the one-dimensional piecewise affine contraction $x \mapsto ax + (1-a)H(T-x)$, the parameter domains for which the mutual inhibitor has permanent oscillations on the diagonal can be explicitly computed [57]. In the threshold space (T_{12}, T_{21}) where a is constant, these domains are collections of squares across the diagonal with some fractal structure and which depend on a. In most cases, the square corresponds to stable periodic oscillations.

When the threshold are close but outside these squares, there are no permanent oscillations. The system however keeps some oscillatory features. Indeed for open sets of initial conditions, the corresponding orbits have some transient oscillations before converging to one of the fixed point $(0,1)$ or $(1,0)$.

The presence of oscillations in the CMN can be attributed to a consequence of delays. The effect of say x_1 crossing the threshold T_{21} at some time t has an influence on x_2 at time $t+1$, after a delay of duration 1. This delay accounts for finite reaction times and finite propagation velocity through the network. During the interval between t and $t+1$, x_2 keeps having the same evolution it had before t. In particular, in spite of x_1 activating (resp. inhibiting) x_2 from t on, this influence being negligible at the origin, the concentration x_2

keeps decreasing (resp. increasing) between t and $t+1$. The activation (resp. inhibition) will only affect x_2 after $t+1$.

This suggests that open sets of trajectories may cross two discontinuities in the interval between t and $t+1$ and this may lead to oscillations. Depending on parameters, the oscillations may or may not be permanent.

The analysis of the CMN shows that permanent oscillations may occur in a two-gene mutual inhibitor provided that both thresholds and both degradation rates are close to each other. As far as application of models to concrete situations is concerned, we point out that experiments on mutual-inhibitor have not revealed permanent oscillations [38]. We suspect this absence of oscillations to be due to important difference between thresholds and/or between degradation rates and we believe such oscillations are possible in systems where both genes have similar parameters.

Of course permanent oscillations could also be exhibited by delay differential equations. However, as dynamical systems with infinite-dimensional phase space[4], their analysis happens to be more involved. The advantage of discrete time models is to mimic the delays by keeping the phase space dimension unchanged.

The difference between discrete time behaviour and that of the differential equation contrasts with classical results on discretization schemes employed in numerical integration of differential equations. Indeed with an appropriate change of time scale (and variable), the CMN can be obtained from the equation in Fig. 9 (a) by applying a finite-difference scheme. However, matching both dynamics requires some conditions on parameters, notably that the discretization step be small enough, depending on the trajectory.

5 Discussion

The functioning and development of living organisms – from bacteria to humans – are controlled by genetic regulatory networks composed of interactions between DNA, RNA, proteins, and small molecules. The size and complexity of these networks make an intuitive understanding of their dynamics difficult to attain. In order to predict the behavior of regulatory systems in a systematic way, we need modeling and simulation tools with a solid foundation in mathematics and computer science. In this chapter, we have examined the modeling and simulation of genetic regulatory networks by means of different types of differential and difference equations.

In any of the continuous time or discrete time version of genetic regulatory network models, due to contraction, the trajectories in every atom converges to a steady state which may or may not belong to the atom. In the second case,

[4] The phase space of a delay differential equation is a set of trajectories prescribed on the delay interval, i.e. a set of functions $x(t)$ with $t \in [T, T+d)$ if d is the delay.

the trajectory eventually leaves the atom and is further deviated. In order to realize the difference between continuous time and discrete time models, one may connect consecutive iterates of CMN by straight lines and consider the whole broken line as a continuous time trajectory.

In the differential equation, when the trajectory crosses a threshold it is instantaneously affected (see Fig. 10 (d)). In the CMN however the trajectory remains unaffected for a while, until the next iteration point is reached. As mentioned before, this difference in dynamics can be interpreted as a delay effect.

To a broader context, modeling the time evolution of physical systems by dynamical systems with either discrete or continuous time has a long history [60]. In the case of spatially extended systems or for systems with many interacting units, the dynamics often involves a coupling parameter.

As seen in other lectures in this school, insights on the influence of interactions can be gained by considering weak couplings. A basic result there is the preservation of the complete phase portrait when passing from an uncoupled to the corresponding weakly coupled hyperbolic system. Mathematically speaking, continuation arguments are applied to prove that a conjugacy holds between both systems.

As piecewise contractions, continuation arguments can also be applied to models of genetic regulatory networks. This allows to mathematically confirm some modularity of the large networks dynamics, a modularity announced in [62]. Indeed, the dynamics of networks composed of weakly interacting circuits (i.e. assuming that the weights K_{ij} of interactions between circuits are small) have the same dynamics as the collection of uncoupled circuits. This argument generally holds for very weakly interacting circuits. However, if their attractor consist of periodic orbit with short period, it may extends to fairly large interaction weights.

Structural stability of weakly interacting systems is probably the unique common feature between models of genetic regulator networks and other systems considered in this school and there are important difference.

The first main difference is the number of nodes to be considered in the network. Indeed whereas most "realistic" regulatory networks typically have from 10 to 100 nodes, many spatially extended systems are reasonably represents macroscopic networks of microscopic interacting units, so a typical size is 10^{23} nodes. A simplifying assumption there is to consider the limit of an infinite number of nodes, the *thermodynamic limit* [55]. It is not clear at all how result on the dynamics at the thermodynamic limit could be extend to intermediate-sized systems, like most genetic regulatory networks. In particular, it would be interesting to know whether the localized, but specific oscillations described in the last section are a more general feature of such systems and whether they are experimentally observed. It is worth noticing that the study of such mesoscopic systems recently became an important challenge in Physics.

In spite of involving a few number of genes, regulatory network usually have complex interaction graphs. Many genes are connected to each other and the incoming and outgoing interactions highly depend on the gene. This is another major difference with spatially extended systems where an usual assumption is that the incoming and outgoing interaction do not depend on the node (translation invariance). As in some neural networks models, simplifying assumptions consists in assuming all-to-all interactions (globally coupled systems) (see the lectures by Maistrenko and by Ermentrout in this school) or random interactions. The assumption seem not to fit exactly with interaction graphs constructed from biological data analysis. Specific graph models and mathematical tools for the analysis of such complex networks of intermediate size are still largely missing.

Another challenge consists in the integration of genetic regulatory networks with metabolic networks, signal transduction networks, and other interaction networks. Even if, for certain problems, the study of one type of network in isolation may be satisfactory, comprehending the functioning of an entire cell obliges us to build models combining gene regulation with metabolism, signal transduction, and other processes. Excellent mathematical models of the individual cellular processes exist nowadays. However, the integration of models of the different processes remains a difficult task. The networks involve different types of interaction, modeled by different types of equation. Moreover, the processes that are concerned evolve on different time-scales, sometimes differing by several orders of magnitude. Among other things, this raises mathematical problems associated with the stiffness of the resulting differential equations. Several approaches for the integration of different types of network have been proposed, but there can be no doubt that the subject remains largely unexplored.

Acknowledgments

The authors would like to thank Johannes Geiselmann for his contributions to this chapter. Part of this work is based on a chapter that will appear in the book *Genomic Signal Processing and Statistics*, published by EURASIP. They also thank Ricardo Coutinho, Bastien Fernandez, Arnaud Meyroneinc, and Edgardo Ugalde for the authorization to use some of their results prior to publication.

References

1. D.J. Lockhart and E.A. Winzeler, Genomics, gene expression and DNA arrays, Nature **405** no. 6788, 827–836 (2000). 307
2. A. Pandey and M. Mann, Proteomics to study genes and genomes, Nature **405** no. 6788, 837–846 (2000). 307

3. R. Alur, C. Belta, F. Ivančíc, V. Kumar, M. Mintz, G.J. Pappas, H. Rubin and J. Schlug, Hybrid modeling and simulation of biomolecular networks, pp. 19–32, in "Hybrid Systems: Computation and Control" (HSCC 2001), M.D. Di Benedetto and A. Sangiovanni-Vincentelli (eds), Lecture Notes in Computer Science **2034**, Springer, 2001. 326
4. R. Ghosh and C.J. Tomlin, Lateral inhibition through Delta-Notch signaling: A piecewise affine hybrid model, pp. 232–246, in "Hybrid Systems: Computation and Control" (HSCC 2001), M.D. Di Benedetto and A. Sangiovanni-Vincentelli (eds), Lecture Notes in Computer Science **2034**, Springer, 2001. 326
5. T. Ideker and T. Galitski and L. Hood, A new approach to decoding life: Systems biology, Annual Review of Genomics and Human Genetics **2**, 343–372, 2001. 307
6. H. Kitano, Systems biology: A brief overview, Science **295**, no. 5560, pp. 1662-1664 (2002). 307
7. J.-C. Leloup and A. Goldbeter, A model for the circadian rhythms in *Drosophila* incorporating the formation of a complex between the PER and TIM proteins, Journal of Biological Rhythms **13**, no. 1, pp. 70–87 (1998). 317
8. H.H. McAdams and L. Shapiro, Circuit simulation of genetic networks, Science **269**, no. 5224, pp. 650–656, 1995. 317
9. B. Lewin, Genes VII, Oxford University Press, Oxford, 1999. 309
10. B.C. Goodwin, Temporal Organization in Cells, Academic Press, N.Y., 1963. 308
11. E.P. van Someren, L.F.A. Wessels and M.J.T. Reinders, Linear modeling of genetic networks from experimental data, pp. 355–366, in Proceedings of the Eight International Conference on Intelligent Systems for Molecular Biology, ISMB 2000, R. Altman, T.L. Bailey, P. Bourne, M. Gribskov, T. Lengauer, I.N. Shindyalov, L.F. Ten Eyck and H. Weissig (eds), Menlo Park, CA, AAAI Press, 2000. 321
12. P. D'Haeseleer, X. Wen, S. Fuhrman and R. Somogyi, Linear modeling of mRNA expression levels during CNS development and injury, pp. 41–52, in "Pacific Symposium on Biocomputing, PSB'99", vol. 4, R.B. Altman, K. Lauderdale, A.K. Dunker, L. Hunter and T.E. Klein (eds), World Scientific Publishing, Singapore, 1999. 321
13. H. de Jong, Modeling and simulation of genetic regulatory systems: A literature review, Journal of Computational Biology **9**, no. 1, 69–105, 2002. 308
14. A. Gilman and A.P. Arkin, Genetic "code": Representations and dynamical models of genetic components and networks, Annual Review of Genomics and Human Genetics **3** (2002), 341–369. 308
15. J. Hasty, D. McMillen, F. Isaacs and J.J. Collins, Computational studies of gene regulatory networks: *In numero* molecular biology, Nature Review Genetics **2**, no. 4 (2001), 268–279. 308
16. P. Smolen, D.A. Baxter and J.H. Byrne, Modeling transcriptional control in gene networks: Methods, recent results, and future directions, Bulletin of Mathematical Biology **62**, no. 2, 247–292, 2000. 308, 313
17. T.M. Gruber and C.A. Gross, Multiple sigma subunits and the partitioning of bacterial transcription space, Annual Review of Microbiology **57** (2003), 441–466. 310
18. A. Ishihama, Functional modulation of *Escherichia coli* RNA polymerase, Annual Review of Microbiology **54** (2000), 499–518. 310
19. R. Hengge-Aronis, Signal transduction and regulatory mechanisms involved in control of the σ^S (RpoS) subunit of RNA polymerase, Microbiology and Molecular Biology Reviews **66**, no. 3, pp. 373–395, 2002. 310, 311, 312

20. H. Tagami and H. Aiba, Role of CRP in transcription activation at *Escherichia coli lac* promoter: CRP is dispensable after the formation of open complex, Nucleic Acids Research **23**, no. 4, pp. 599-605, 1995. 310
21. K.W. Kohn, Molecular interaction maps as information organizers and simulation guides, Chaos **11**, no. 1, 1–14 (2001). 311
22. F. Repoila, N. Majdalani and S. Gottesman, Small non-coding RNAs, coordinators of adaptation processes in *Escherichia coli*: The RpoS paradigm, Molecular Microbiology **48**, no. 4, 855–862 (2003). 311
23. P. Brazhnik, A. de la Fuente and P. Mendes, Gene networks: How to put the function in genomics, Trends in Biotechnology **20**, no.11, 467–472 (2002). 312, 320
24. J. Monod and F. Jacob, General conclusions: Teleonomic mechanisms in cellular metabolism, growth, and differentiation, pp. 389–401, Cold Spring Harbor Symposium on Quantitative Biology, vol. 26, Cold Spring Harbor, NY, 1961. 313
25. D. Endy, L. You, J. Yin and I.J. Molineux, Computation, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes, Proceedings of the National Academy of Sciences of the USA **97**, no. 10, pp. 5375-5380, 2000. 317
26. R. Thomas and R. d'Ari, Biological Feedback, CRC Press, Boca Raton, FL, 1990. 313, 316, 326
27. T.S. Gardner and C.R. Cantor and J.J. Collins, Construction of a genetic toggle switch in *Escherichia coli*, Nature **403** no. 6767, 339–342 (2000). 316
28. M. Ptashne, A Genetic Switch: Phage λ and Higher Organisms, Cell Press & Blackwell Science, Cambridge, MA, 2nd ed, 1992. 314
29. G. Yagil and E. Yagil, On the relation between effector concentration and the rate of induced enzyme synthesis, Biophysical Journal **11**, no. 1, 11–27, 1971. 314
30. D. Kaplan and L. Glass, Understanding Nonlinear Dynamics, Springer-Verlag, New York, 1995. 314, 316
31. J.D. Lambert, Numerical Methods for Ordinary Differential Equations, Wiley, Chichester, 1991. 316
32. M. Hucka, A. Finney, H.M. Sauro, H. Bolouri, J. Doyle and H. Kitano, The ERATO Systems Biology Workbench: Enabling interactions and exchange between software tools for computational biology, pp. 450–461, in Pacific Symposium on Biocomputing, PSB'02, vol. 7, R.B. Altman and A.K. Dunker and L. Hunter and K. Lauderdale and T.E. Klein (eds), World Scientific Publishing, Singapore, 2002. 317
33. L. Ljung, System Identification: Theory for the User, Prentice Hall, Upper Saddle River, NJ, 2nd ed, 1999. 317, 320
34. G. von Dassow, E. Meir, E.M. Munro and G.M. Odell, The segment polarity network is a robust developmental module, Nature **406**, no. 6792, 188-192, 2000. 317
35. N. Barkai and S. Leibler, Robustness in simple biochemical networks, Nature **387**:913–917, 1997. 317
36. W.J. Rugh, Linear System Theory, Prentice Hall, Englewood Cliffs, NJ, 2nd ed, 1996. 318
37. P. D'haeseleer, S. Liang and R. Somogyi, Genetic network inference: From coexpression clustering to reverse engineering, Bioinformatics **16**, no. 8, 707–726, 2000. 320
38. T.S. Gardner, D. di Bernardo, D. Lorenz and J.J. Collins, Inferring genetic networks and identifying compound mode of action via expression profiling, Science **301**, no. 5629, 102–104, 2003. 320, 321, 334

39. M. Ronen, R. Rosenberg, B.I. Shraiman and U. Alon, Assigning numbers to the arrows: Parameterizing a gene regulation network by using accurate expression kinetics, Proceedings of the National Academy of Sciences of the USA (2002), **99**, no. 16, pp. 10555–10560. 321
40. R. Edwards, H.T. Siegelmann, K. Aziza and L. Glass, Symbolic dynamics and computation in model gene networks, Chaos **11**, no. 1, 160–169, 2001. 321, 322
41. L. Glass and S.A. Kauffman, The logical analysis of continuous non-linear biochemical control networks, Journal of Theoretical Biology **39**, no. 1, 103–129, 1973. 321
42. T. Mestl, E. Plahte and S.W. Omholt, A mathematical framework for describing and analysing gene regulatory networks, Journal of Theoretical Biology **176**, no. 2, 291–300, 1995. 321
43. E.H. Snoussi, Qualitative dynamics of piecewise-linear differential equations: A discrete mapping approach, Dynamics and Stability of Systems **4**, no. 3-4, 189–207, 1989. 321
44. J.-L. Gouzé, T. Sari, A class of piecewise linear differential equations arising in biological models, Dynamical Systems **17**, no. 4, 299–316, 2003. 323
45. H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari and J. Geiselmann, Qualitative simulation of genetic regulatory networks using piecewise-linear models, Bulletin of Mathematical Biology **66**, no. 2, 301–340, 2004. 325, 326
46. H. de Jong, J. Geiselmann, G. Batt, C. Hernandez and M. Page, Qualitative simulation of the initiation of sporulation in *B. subtilis*, Bulletin of Mathematical Biology **66**, no. 2, 261–300, 2004. 326
47. H. de Jong, J. Geiselmann, C. Hernandez and M. Page, Genetic Network Analyzer: Qualitative simulation of genetic regulatory networks, Bioinformatics **19**, no. 3, 336–344, 2003. 326
48. R. Thomas, D. Thieffry and M. Kaufman, Dynamical behaviour of biological regulatory networks: I. Biological role of feedback loops and practical use of the concept of the loop-characteristic state", Bulletin of Mathematical Biology **57**, no. 2, 247–276, 1995. 326
49. L. Mendoza, D. Thieffry and E.R. Alvarez-Buylla, Genetic control of flower morphogenesis in *Arabidopsis thaliana*: A logical analysis, Bioinformatics **15**, no. 7-8, 593-606, 1999. 326
50. L. Sánchez and D. Thieffry, A logical analysis of the *Drosophila* gap genes, Journal of Theoretical Biology **211**, no. 2, 115–141, 2001. 326
51. L. Sánchez and D. Thieffry, Segmenting the fly embryo: A logical analysis of the pair-rule cross-regulatory module, Journal of Theoretical Biology **224**, no. 4, 517–538, 2003. 326
52. D. Thieffry and R. Thomas, Dynamical behaviour of biological networks: II. Immunity control in bacteriophage lambda, Bulletin of Mathematical Biology **57**, no. 2, 277–297, 1995. 326
53. H. de Jong, J. Geiselmann and D. Thieffry, Qualitative modeling and simulation of developmental regulatory networks, pp. 109-134, in On Growth, Form, and Computers, S. Kumar and P.J. Bentley (eds), Academic Press, London, 2003. 326
54. R. Coutinho, Dynâmica simbólica linear, PhD thesis, University of Lisbon (1999). 327, 329
55. D. Volchenkov and R. Lima, Homogeneous and scalable gene expression regulatory networks with random layouts of switching parameters, submitted (q-bio.MN/0311031) (2004). 331, 335

56. R. Coutinho and B. Fernandez, Extended symbolic dynamics in bistable CML: Existence and stability of fronts, Physica D **108** (1997) 60–80. 327
57. R. Coutinho, B. Fernandez, R. Lima and A. Meyroneinc, Discrete time models for the piecewise affine dynamics of genetic regulatory networks, Preprint Centre de Physique Théorique (2004). 327, 329, 332, 333
58. A. Meyroneinc, Modélisation de réseaux de régulation génétique par un modèle CML, mémoire de DEA, Centre de Physique Théorique (2003). 327
59. R. Lima and E. Ugalde, Centre de Physique Théorique. In preparation. 331
60. H. Poincaré, Les méthodes nouvelles de la mécanique céleste, Dover, NY, 1957. 335
61. H. Bolouri and E.H. Davidson, Modeling transcriptional regulatory networks, BioEssays **24**, no. 12, 1118–1129, 2002. 308
62. D. Thieffry and D. Romero, The modularity of biological regulatory networks, BioSystems **50** (1999) 49–59. 335

Desynchronization and Chaos in the Kuramoto Model

Y.L. Maistrenko[1,2,4], O.V. Popovych[1,4], and P.A. Tass[1,3]

[1] Institute of Medicine, Research Centre Jülich, 52425 Jülich, Germany
[2] Central Institute for Electronics, Research Center Jülich, 52425 Jülich, Germany
[3] Department of Stereotaxic and Functional Neurosurgery, University Hospital, 50924 Cologne, Germany
[4] Institute of Mathematics, Academy of Sciences of Ukraine, 01601 Kyiv, Ukraine
maistrenko@imath.kiev.ua, o.popovych@fz-juelich.de, p.tass@fz-juelich.de

Abstract. The Kuramoto model of N globally coupled phase oscillators is an essentially non-linear dynamical system with a rich dynamical behavior and a high relevance for numerous applications. We study the Kuramoto model from the standpoint of bifurcation theory and chaos theory of low-dimensional dynamical systems. We focus on the desynchronization transition and the role of the Cherry flow in it. Furthermore, we study chaos, hyperchaos, and multistability. With an additional symmetry condition the Kuramoto model is reduced to the Winfree type model of half the dimension. We find out that the dynamics in the symmetric manifold is responsible for the desynchronization transition. With a further decrease of the coupling, the manifold loses its transverse stability, which gives rise to a highly developed hyperchaotic behavior.

1 Introduction

1.1 Goal

Starting with the work of Winfree [2] and Kuramoto [3], there has been a growing interest in synchronization of globally coupled limit cycle oscillators [4, 5, 6, 7, 8, 9]. Possible applications include many self-organizing systems in physics, chemistry, biology, and medicine and range, e.g., from Josephson junction arrays [10], semiconductor lasers arrays [11], chemistry [12], cardiac pacemaker cells [13], flashing fireflies [14] to the development of demand-controlled brain pacemakers for the therapy of neurological and psychiatric diseases [16, 17]. Fundamental to all such applications is, to understand mechanisms that cause synchronization or desynchronization.

If the coupling between limit cycle oscillators is strong enough, in general phase synchronization occurs: All oscillators start to rotate with the same

Y.L. Maistrenko, O.V. Popovych and P.A. Tass: *Desynchronization and Chaos in the Kuramoto Model*, Lect. Notes Phys. **671**, 285–306 (2005)
www.springerlink.com

average frequency. With a decrease of the coupling strength, a desynchronization occurs: The oscillators split into groups of different average frequencies, such that inside each group the frequency is the same. This transition is called a *frequency-splitting bifurcation*. It represents a common property of very different ensembles of coupled oscillators with both local and global coupling, characterized by regular or chaotic dynamics (see e.g. [18, 19, 20, 21] and references therein).

General properties of the phase dynamics in ensembles of nonlinear limit cycle oscillators with weak global coupling are described by the Kuramoto model [3]

$$\dot{\psi}_i = \omega_i + \frac{K}{N}\sum_{j=1}^{N}\sin(\psi_j - \psi_i)\,, \qquad i = 1, \ldots, N \tag{1}$$

where ψ_i are phase variables ω_i are natural frequencies, $K > 0$ is a constant coupling parameter. In the thermodynamic limit $N \to \infty$ a transition from the synchronized state to a complete desynchronization occurs, when the strength of the coupling decreases below a certain critical value [3].

In spite of numerous studies, in the finite-dimensional Kuramoto model the desynchronization mechanism is still far from being well understood (see, e.g., [19] and references therein). This issue is the starting point of the present chapter. Unlike Kuramoto, Strogatz and others [3, 19] we do not use well-developed statistical techniques. In contrast, our goal is to study the Kuramoto model (1) from the point of view of bifurcation theory and chaos theory of low-dimensional dynamical systems.

By introducing new variables, the phase differences $\varphi_i = \psi_{i+1} - \psi_i$, the dimension of the Kuramoto model (1) is reduced by one. The system's dynamics is then given by a smooth flow on an $(N-1)$-dimensional torus $\mathbb{T}^{N-1}$.

The simplest, non-trivial example is given by the system of $N = 3$ Kuramoto equations. In this case, the reduced dynamics of the phase differences acts on a two-dimensional torus $\mathbb{T}^2$. We consider this example in detail in Sect. 2.2 to demonstrate the mechanism of the frequency-splitting bifurcation [22]. The complete bifurcation diagram is plotted, and the existence of a Cherry flow is shown for a rather large region of the parameter plane. The Cherry flow exists before the desynchronization transition and controls the course of the frequency-splitting bifurcation. We conclude that, depending on the parameters, two robust bifurcation scenarios can be realized causing an instant emergence of two or three different average frequencies, respectively.

For dimension $N \geq 4$, the behavior of the Kuramoto model can be chaotic. For $N = 4$, we demonstrate the existence of a chaotic attractor in the three-dimensional system phase space. The chaotic attractor arises due to a destruction of a two-dimensional resonant torus, which is in agreement with the known torus-destruction scenario for the transition to chaos [23]. The chaotic attractor exists for a wide range of the coupling strength. Finally, it disappears in a boundary crisis. For higher dimensions, the dynamics of the Kuramoto model can be hyperchaotic. We demonstrate this phenomenon in the case of

$N = 7$ by calculating all six Lyapunov exponents. Two of them turn out to be positive for a large interval of the coupling strength (see Sect. 3.5).

In Sect. 3 we consider a class of symmetric Kuramoto models. They are characterized by the existence of an $[N/2]$-dimensional invariant manifold which represents a hyperplane in the whole N-dimensional phase space. Our findings show that, the symmetric invariant manifold is stable for an essential part of the coupling parameter range. Moreover, the desynchronization transition which is a main subject of our study, takes place in the system when the dynamics is restricted to the invariant manifold.

1.2 The Winfree Model

Phase synchronization in ensembles of globally coupled oscillators was first studied by Winfree [1, 2] who proposed the model

$$\dot{\psi}_i = \omega_i + \frac{K}{N}\, R(\psi_i) \sum_{j=1}^{N} P(\psi_j), \quad i = 1, \ldots, N\,. \tag{2}$$

According to the Winfree model (2), dynamics of each phase ψ_i reacts on its own state through the function R and is influenced by a mean-field type function P. A special case of the Winfree model with $R = -\sin\psi$ and $P = 1 + \cos\psi$ was considered in [7].

1.3 Kuramoto Model

Kuramoto found another model describing the evolution of the phase dynamics of an ensemble of N weakly coupled, nearly identical limit cycle oscillators. He showed that, under these assumptions, the long-term dynamics is given by the following general system

$$\dot{\psi}_i = \omega_i + \sum_{j=1}^{N} \Gamma_{ij}(\psi_j - \psi_i)\,, \qquad i = 1, \ldots, N \tag{3}$$

where Γ_{ij} is an interaction function of phase differences. The dynamics of system (3) is determined by the form of the function Γ_{ij}. Kuramoto considered the simplest, non-trivial case of equally weighted, globally coupled oscillators: $\Gamma_{ij}(\varphi) = KN^{-1}\sin\varphi$ for all i and j, with constant coupling strength K. System (3) with this coupling term takes the form of model (1) and is known as the Kuramoto model.

For the statistical theory of the Kuramoto model (1), well developed by Kuramoto, Strogatz and others (see [19] for an excellent survey), the natural frequencies ω_i are assumed to be given by a probability density function $g(\omega)$. The function $g(\omega)$ is often assumed to be unimodal and symmetric with respect to a mean frequency Ω, i.e. $g(\Omega + \omega) = g(\Omega - \omega)$ for all ω. g can, e.g., be a Gaussian, a Lorentzian or a uniform distribution.

The form of (3) is invariant with respect to a change of variables given by $\psi_i \mapsto \psi_i + \Omega t$, which reduces the mean frequency to zero by choosing a coordinate system rotating with frequency Ω. New natural frequencies are obtained by the subtraction $\omega_i - \Omega$. Then, in the new variables, the natural frequencies ω_i are symmetrically distributed around zero: $g(\omega) = g(-\omega)$.

Following Kuramoto, we consider the phase variables $\psi_1, \psi_2, \ldots, \psi_N$ of (1) as a set of points moving along the unit circle in the complex plane. An important characteristic of the collective dynamics of the ensemble is given by the complex order parameter

$$r(t)\ \mathrm{e}^{i\Psi(t)} = \frac{1}{N}\sum_{j=1}^{N} \mathrm{e}^{i\psi_j(t)} \tag{4}$$

which is the circular mean of the circular phases $\exp[i\psi_j(t)]$. The amplitude $r(t)$ of the order parameter measures the extent of in-phase synchronization of the ensemble, and $\Psi(t)$ is the average phase of the ensemble. $r(t) = 1$ corresponds to complete in-phase synchronization, whereas $r(t) = 0$ corresponds to an absence of in-phase synchronization. Note that the case $r(t) = 0$ may be associated with complete desynchronization (in terms of a uniform distribution of the phases) or with cluster states, e.g., two symmetric clusters synchronized in anti-phase (see, e.g., [16]).

Using the complex order parameter (4), the Kuramoto model becomes

$$\dot{\psi}_i = \omega_i - Kr\sin(\psi_i - \Psi), \qquad i = 1, \ldots, N\ . \tag{5}$$

Under this representation the role of the coherence term $r(t)$ becomes evident. $r(t)$ acts as an effective coupling strength. By providing a positive feedback between coupling and coherence, $r(t)$ contributes to a circular causality underlying the emergence of self-organized synchronized states.

Kuramoto has studied the case where the amplitude $r(t)$ of the order parameter is constant and the average phase $\Psi(t)$ is rotating with a constant frequency Ω (clearly, Ω can be reduced to zero).

In our study presented below in Sect. 3, we consider the N-dimensional Kuramoto model fulfilling a symmetry condition for the natural frequency distribution. In the corresponding invariant manifold M (which represents a $[N/2]$-dimensional hyperplane in the whole phase space) the average phase $\Psi(t)$ vanishes, and (5) turns into a Winfree type model (2), such that the amplitude r of the order parameter plays role of the mean-field term P.

1.4 More General Kuramoto Models

For the standard Kuramoto model (1) the coupling function reads $\Gamma = KN^{-1}\sin\varphi$. More generally, the coupling function $\Gamma_{ij}(\varphi)$ should contain terms that involve not only a simple sine term, but also higher harmonics. Tass [15, 16] considered the four-mode model with

$$\Gamma(\varphi) = \frac{1}{N} \sum_{m=1}^{4} [K_m \sin(m\varphi) + C_m \cos(m\varphi)]$$

and stressed the importance of higher harmonics for the emergence of cluster states. Furthermore, he extended this model by incorporating stimulation terms in order to develop desynchronizing stimulation techniques (see below) [17].

Hansel, Mato and Meunier [24], and later Kori and Kuramoto [25] examined the two-mode case with a phase shift between the harmonics:

$$\Gamma(\varphi) = \frac{K}{N} [\sin(\varphi - \alpha) - c \sin 2\varphi] \ .$$

The phase shift α plays role of a delay and controls the competition between the two harmonics in Γ. While the first term in Γ promotes an in-phase synchronization, the second term causes an anti-phase synchronization. For a particular choice of the parameter values, $\alpha = 1.25$ and $c = 0.25$, an interesting phenomenon occurs, the so-called slow switching, which was explained in the terms of a heteroclinic loop connecting a pair of cluster states.

The Kuramoto model with a time delay can be written as

$$\dot{\psi}_i = \omega_i + \frac{K}{N} \sum_{j=1}^{N} \sin[\psi_j(t - \tau) - \psi_i(t)] \ , \qquad i = 1, \ldots, N \ . \tag{6}$$

Schuster and Wagner [26] considered the simplest, but definitely non-trivial case of only $N = 2$ coupled phase oscillators in (6). They found that the delay causes the appearance of new stable phase-locked states with different frequencies Ω_i, where their number grows with an increase of τ. This property of multistability is a major difference between the time-delayed system (6) and the standard ordinary differential Kuramoto model (1). In the standard Kuramoto model (1) with sine coupling only one, just the mean frequency Ω can be stabilized in the synchronized state. See [27] for more studies of the multistability phenomenon and other properties of the time-delayed Kuramoto model.

1.5 Kuramoto Model and Deep Brain Stimulation

In several neurological diseases brain function is severely perturbed by pathologically enhanced synchronization. For example, resting tremor in Parkinson's disease (PD) appears to be caused by a cluster of neurons located in the thalamus and the basal ganglia which fire synchronously at a frequency similar to that of the tremor (3–6 Hz). While under physiological conditions these neurons fire incoherently, in patients with PD this cluster acts like a pacemaker and activates cortical areas, which causes the peripheral shaking [28]. In patients with advanced PD or with essential tremor who do not respond to drug therapy, depth electrodes are chronically implanted in target

areas like the thalamic ventralis intermedius nucleus or the subthalamic nucleus [29]. As yet, electrical deep brain stimulation (DBS) is performed by administering a permanent high-frequency (>100 Hz) periodic pulse train via the depth electrodes. DBS mimics the effect of tissue lesioning and appears to suppress the activity of the pacemaker-like cluster which, in turn, suppresses the peripheral tremor [29]. DBS has been developed empirically, mainly based on observations during stereotaxic neurosurgery. The advent of nonlinear dynamics and statistical physics in the field of clinically oriented neuroscience lead to a model-based development of stimulation techniques [16, 17].

For this, in a first step the impact of stimulation on an ensemble of uncoupled oscillators in the presence of noise has been studied [30]. In this approach a phase oscillator has been used as a simple model for an oscillatory neuron. Next, in addition the oscillators' mutual coupling has been taken into account [16], which led to the model

$$\dot{\psi}_j = \omega_j + \sum_{k=1}^{N} \Gamma(\psi_j - \psi_k) + X(t)\, S(\psi_j) + F_j(t) \tag{7}$$

where ψ_j is the phase of the jth phase oscillator, Γ is a 2π-periodic global coupling, S is a 2π-periodic, time independent function modelling the stimulus, where $X(t) = 1$ during stimulation, and $X(t) = 0$ without stimulation [16]. The random forces $F_j(t)$ are modelled by Gaussian white noise fulfilling $\langle F_j(t)\rangle = 0$, $\langle F_j(t)\ F_k(t')\rangle = D\delta_{jk}\delta(t-t')$ with constant noise amplitude D.

Model (7) formed the basis for the study of transient, stimulation-induced dynamics. In particular, such mechanisms have been used to develop stimulation techniques which provide effective means for demand-controlled desynchronization [17].

These stimulation techniques have recently been verified in simulations of physiologically realistic neural network models for target populations relevant for DBS. It should be mentioned, that not only the steady state dynamics of the Kuramoto model, but also its repertoire of transient dynamics appears to approximate a large class of steady state and transient network dynamics in a generic way. It is advantageous to develop stimulation techniques by using the generic model (7) and testing them in microscopic neural network models, instead of exclusively using the high-dimensional microscopic models with their huge number of model parameters, which are often not accessible to analytical or sound numerical analysis. Furthermore, in the meantime the demand-controlled stimulation techniques have been verified in experiments in both patients and animals. The corresponding papers are just being submitted.

2 Frequency-Splitting Bifurcation

2.1 Simplest Example: $N = 2$

Let us consider the case of only two coupled phase oscillators, which is given by the Kuramoto model with $N = 2$:

$$\dot{\psi}_1 = \omega_1 + (K/2)\ \sin(\psi_2 - \psi_1)$$
$$\dot{\psi}_2 = \omega_2 + (K/2)\ \sin(\psi_1 - \psi_2)\ .$$

Two phases ψ_1 and ψ_2 are called synchronized if their difference $\varphi(t) = \psi_2(t) - \psi_1(t)$ is bounded for all $t > 0$. For the phase difference $\varphi(t)$ one immediately obtains

$$\dot{\varphi} = \Delta - K \sin \varphi \tag{8}$$

with $\Delta = \omega_2 - \omega_1$. The phase difference $\varphi(t)$ is bounded, if (8) has equilibria, i.e. fixed points. Desynchronization occurs when the fixed points disappear.

There are two fixed points, $\phi_{(s)} = \arcsin(\frac{\Delta}{K})$ which is stable, and $\phi_{(u)} = \pi - \arcsin(\frac{\Delta}{K})$ which is unstable. They exist if and only if $K \geq K_c$, where

$$K_c = \Delta$$

is a critical bifurcation value. For any $K > K_c$, both phase variables ψ_1 and ψ_2 rotate with the same average frequency, the mean frequency $\Omega = (\omega_1 + \omega_2)/2$. A bifurcation occurs at $K = K_c$: for $K < K_c$ the phases ψ_1 and ψ_2 rotate with different average frequencies $\overline{\omega}_1$ and $\overline{\omega}_2$, where

$$\overline{\omega}_{1,2} = \Omega \pm \pi \left(\int_0^{2\pi} \frac{d\varphi}{\Delta - K \sin(\varphi)} \right)^{-1} .$$

The desynchronization transition at $K = K_c$ is called a cluster-splitting bifurcation. Beyond the bifurcation, as K tends to zero, the average frequencies $\overline{\omega}_1$ and $\overline{\omega}_2$ approach the natural frequencies ω_1 and ω_2. A typical bifurcation diagram is presented in Fig. 1.

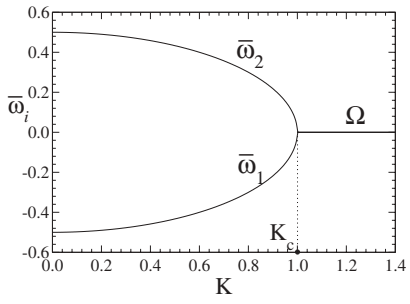


Fig. 1. Frequency-splitting bifurcation diagram for the Kuramoto model from (1) with $N = 2$ phase oscillators with the natural frequencies ω_i: $\{-0.5, 0.5\}$. $\overline{\omega}_i = \langle \dot{\psi}_i \rangle$, where $\langle \cdot \rangle$ denotes averaging over time. Ω is the mean frequency

2.2 From Cherry Flow Scenario to Frequency-Splitting: $N = 3$

Next, we consider the Kuramoto model of $N = 3$ phase oscillators given by

$$\begin{aligned}\dot{\psi}_1 &= \omega_1 + (K/3)\,[\sin(\psi_2 - \psi_1) + \sin(\psi_3 - \psi_1)]\\ \dot{\psi}_2 &= \omega_2 + (K/3)\,[\sin(\psi_1 - \psi_2) + \sin(\psi_3 - \psi_2)]\\ \dot{\psi}_3 &= \omega_3 + (K/3)\,[\sin(\psi_1 - \psi_3) + \sin(\psi_2 - \psi_3)]\ .\end{aligned}$$

As for $N = 2$, there exists a critical bifurcation value $K = K_c$ for the splitting of the common average frequency $\overline{\omega}_1 = \overline{\omega}_2 = \overline{\omega}_3 = \Omega$. The course of the splitting depends on the ratio of the differences of the natural frequencies ω_1, ω_2 and ω_3 and may produce two or three different frequencies.

Figure 2 illustrates three different types of the frequency-splitting bifurcations. In Fig. 2(a) a standard bifurcation sequence is plotted. With decreasing K, first, at $K = K_c$, complete synchronization $\overline{\omega}_1 = \overline{\omega}_2 = \overline{\omega}_3 = \Omega$ turns into two unequal frequencies $\overline{\omega}_1 = \overline{\omega}_2$ and $\overline{\omega}_3$. Then, at a smaller coupling parameter value K_{c1}, the first cluster (with $\overline{\omega}_1 = \overline{\omega}_2$) is also splitted, causing a complete desynchronization: $\overline{\omega}_1 \neq \overline{\omega}_2 \neq \overline{\omega}_3$. Both bifurcations may be referred to as cluster-doubling bifurcations. In Fig. 2(b) and Fig. 2(c) two examples of a cluster-tripling bifurcation are presented. The synchronous motion breaks up at $K = K_c$, instantly producing three different frequencies. Furthermore, just after the bifurcation, successive frequency differences $\overline{\Delta}_1 = \overline{\omega}_2 - \overline{\omega}_1$ and $\overline{\Delta}_2 = \overline{\omega}_3 - \overline{\omega}_2$ become locked in a 1:1 resonance [Fig. 2(b)] and in a 1:2 resonance [Fig. 2(c)]: $\overline{\Delta}_1/\overline{\Delta}_2 = 1:1$ and $\overline{\Delta}_1/\overline{\Delta}_2 = 1:2$, respectively. Below we will show, that, depending on the parameters, at the bifurcation any other $p:q$ resonance can robustly be obtained.

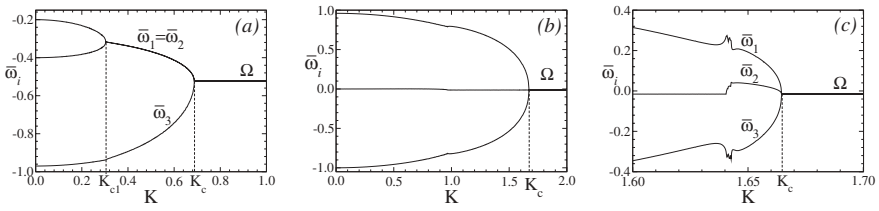


Fig. 2. Frequency-splitting bifurcation diagrams for the Kuramoto model from (1) with $N = 3$ phase oscillators, with a non-uniform distribution of the natural frequencies ω_i: (**a**) $\{-0.97, -0.4, -0.2\}$, (**b**) $\{-1.0, 0.0, 0.96\}$, and (**c**) $\{-1.0, 0.0, 0.953\}$. $\overline{\omega}_i = \langle\dot{\psi}_i\rangle$, where $\langle\cdot\rangle$ denotes averaging over time. Ω is the mean frequency

The mechanism of the frequency-splitting bifurcation can be explained by considering the corresponding two-dimensional system of the phase differences $\varphi_1 = \psi_2 - \psi_1$ and $\varphi_2 = \psi_3 - \psi_2$:

$$\begin{aligned}\dot{\varphi}_1 &= \Delta_1 + (K/3)\,[\sin(\varphi_2) - \sin(\varphi_1 + \varphi_2) - 2\sin(\varphi_1)]\\ \dot{\varphi}_2 &= \Delta_2 + (K/3)\,[\sin(\varphi_1) - \sin(\varphi_1 + \varphi_2) - 2\sin(\varphi_2)]\end{aligned} \tag{9}$$

where $\Delta_1 = \omega_2 - \omega_1$ and $\Delta_2 = \omega_3 - \omega_2$. We consider the flow Φ defined by (9), which is a flow on a two-dimensional torus $\mathbb{T}^2 = [0, 2\pi[^2$.

Let first $\Delta_1 = \Delta_2 = 0$. In this case, the phase portrait of the system (9) does not depend on $K > 0$ and synchronization always takes place, see Fig. 3(a). There are six equilibria each characterized by two Lyapunov exponents λ_1 and λ_2, namely: sink at the origin $O(0,0)$, $\lambda_1 = \lambda_2 = -K$; three saddles $S_1(0,\pi), S_2(\pi,0)$, $S_3(\pi,\pi)$, $\lambda_1 = K$, $\lambda_2 = -K/3$; and two sources $N_1(2\pi/3, 2\pi/3)$, $N_2(4\pi/3, 4\pi/3)$, $\lambda_1 = \lambda_2 = K/2$. Any trajectory of system (9) originating anywhere (except for the unstable equilibria and the stable manifolds of the saddles) approaches the stable node O without even a single rotation around the torus.

Next, we consider the case with $\Delta_1 = \Delta_2 \stackrel{def}{=} \Delta > 0$. Under this condition, system (9) preserves its symmetry with respect to the diagonal $\{\varphi_1 = \varphi_2\}$ which is an invariant manifold – we denote in by M – of the torus flow Φ. Equation (9) on the manifold M turn into

$$\dot{\varphi} = \Delta - \frac{K}{3}[\sin(\varphi) + \sin(2\varphi)] \tag{10}$$

where $\varphi \stackrel{def}{=} \varphi_1 = \varphi_2$. For the values of Δ, that are positive and small enough, (9) has six equilibria O, N_1, N_2, S_1, S_2, S_3 (see Fig. 3a). With decreasing control parameter K/Δ, first, S_2 and N_2 annihilate in an inverse saddle-node bifurcation. Then, the saddles S_1 and S_3 collide with the unstable node N_1, in this way producing, instead, a saddle S. This is a combination of two bifurcations, local (inverse pitchfork) and global (homoclinic). After this bifurcation, we obtain a phase portrait as shown in Fig. 3(b). It contains two equilibria, a stable node O and a saddle S, and an unstable periodic orbit P (born in the homoclinic bifircation). This is a so-called Cherry flow

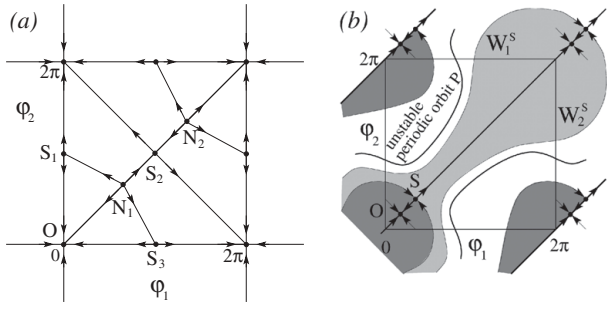


Fig. 3. Schematic representation of the phase portraits for (9) with $\Delta_1 = \Delta_2 = \Delta$. (**a**) $\Delta = 0$: The phase portrait contains a stable node O, saddles S_1, S_2, and S_3, and unstable nodes N_1 and N_2. (**b**) $\Delta > 0$ and $1.70 < K/\Delta < 3$ (Cherry flow): The phase portrait contains a stable node O, a saddle S, and an unstable periodic orbit P, and is foliated by stable manifolds W_1^S and W_2^S of the saddle S, as indicated by *gray* regions

[31, 32, 33, 34, 35]. Following [34], a *Cherry flow* is defined as a smooth flow on a two-dimensional torus, which has two equilibria, one being a saddle and the other being a sink or source, and it has rotating trajectories of some rotation number which is called *winding ratio*. In the symmetric case $\Delta_1 = \Delta_2 = \Delta > 0$, obviously, the winding ratio of the Cherry flow equals 1:1.

With a further decrease of the coupling strength the frequency-splitting bifurcation occurs at $K = K_c \approx 1.70\Delta$. The stable node O and the saddle S vanish in a saddle-node bifurcation producing a stable periodic orbit γ (which coincides with the torus diagonal M). As a result, both phase differences $\varphi_1 = \psi_2 - \psi_1$ and $\varphi_2 = \psi_3 - \psi_2$ start to grow, which corresponds to a desynchronization in the original Kuramoto system in form of a cluster tripling.

Let now $\Delta_1 \neq \Delta_2$ and let, for definiteness, $\Delta_1 < \Delta_2$. In the bifurcation diagram in Fig. 4(a) three saddle-node bifurcation curves B_1, B_2, and B_c, are shown. They correspond to pairwise annihilations of the six equilibria of (9) as follows: N_2 meets S_2 at curve B_1, N_1 meets S_3 at B_2, and, finally, O meets S_1 at B_c. Above B_2, the flow Φ has no circular rotations around the torus. The Cherry flow exists in the hatched parameter region between the B_2 and B_c

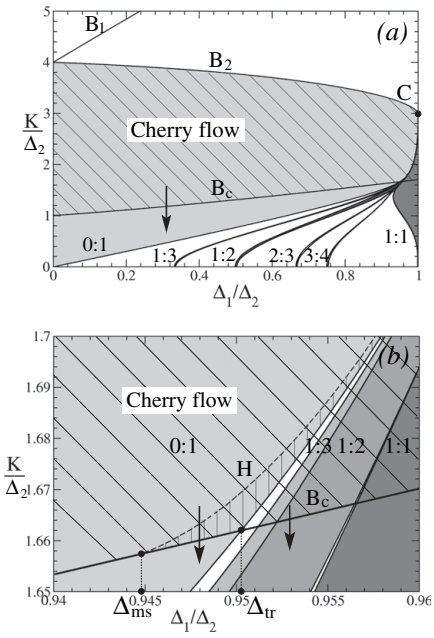


Fig. 4. (**a**) Two-parameter bifurcation diagram for (1) with N = 3. The Cherry flow region is hatched. The main resonant tongues 0:1, 1:3, 1:2, 2:3, 3:4, and 1:1 are shown. (**b**) Enlargement from (**a**). Parameter variations along the *arrow* in (**a**) and the *right arrow* in (**b**) correspond to the bifurcation diagrams in Figs. 2(**a**) and 2(**b**), respectively

curves. Its winding ratio varies continuously with the parameters as a devil's staircase, and may be rational or irrational. In Fig. 4(a) two main resonant tongues are shown, with winding ratio 0:1 (light gray) and 1:1 (dark gray). Other resonant tongues, with winding ratios $p : q$, are aligned in between the main 0:1 and 1:1 tongues. All the tongues extend up to the singular parameter point $C = (1, 3)$, where they glue together and end. (Quasiperiodic torus rotations take place for the fractal parameter set which complements the union of the resonant tongues and is very thin: both its Lebesgue measure and its Hausdorff dimension equal zero; see [32, 33]).

Figure 4(b) presents an enlargement of the bifurcation diagram shown in Fig. 4(a) for the region where the B_c bifurcation curve intersects the resonant tongues. The enlargement shows that the resonant tongues, which exist above B_c, are naturally continued below B_c, in the desynchronization region. The frequency-splitting bifurcation occurs when the parameter point crosses B_c.

For all $\Delta_1/\Delta_2 < \Delta_{tr}$, where Δ_{tr} is the intersection point of the 0:1 tongue boundary with the B_c bifurcation curve, a cluster doubling takes place: Only the second phase difference φ_2 grows. At other parameter values, i.e. for $\Delta_1/\Delta_2 > \Delta_{tr}$, the desynchronization transition always shows up as a cluster-tripling: Both phase differences φ_1 and φ_2 start growing, in this way causing an instant splitting of all three phase variables ψ_1, ψ_2, and ψ_3. The transitions are illustrated in Fig. 2(b) and Fig. 2(c) respectively.

The cluster-splitting desynchronization transition shown is more complicated near boundaries of the Cherry flow resonant tongues. Indeed, the resonant tongues contain thin boundary layer strips, where a stable periodic orbit exists (see [35]). The periodic orbit is born in a homoclinic bifurcation of the saddle S. One of the homoclinic bifurcation curves is shown in Fig. 4(b) as a dashed line H inside the 0:1 tongue. Hence, two attracting states co-exist in the vertically hatched region between the curve H and the tongue boundary: (1) the Cherry flow steady state O, and (2) the stable periodic orbit emerging in the homoclinic bifurcation at H. The multistability region is attached to the B_c saddle-node curve, comprising a non-empty interval $\Delta_{ms} < \Delta_1/\Delta_2 < \Delta_{tr}$. If a desynchronization transition takes place in this interval, e.g., when K/Δ_2 is decreased along the left arrow in Fig. 4(b), synchronized and desynchronized states co-exist for the range of parameters between H and B_c. This sideband type multistability regions exist in any Cherry flow resonant tongue (not shown in Fig. 4(b) being very thin). They extend along the tongues' boundaries up to the singular parameter point $C(1, 3)$. We conclude that in the Kuramoto model a stable desynchronized state can, actually, arise even far before the saddle-node disappearance of the phase-locked state.

2.3 Multiple Frequency-Splitting: $N = 5$

Figures 5(a) and 5(b) illustrate two different types of the cluster-splitting bifurcation in the five-dimensional Kuramoto model. Natural frequencies ω_i ($i = \overline{1,5}$) are randomly chosen close to a uniform distribution. As one can

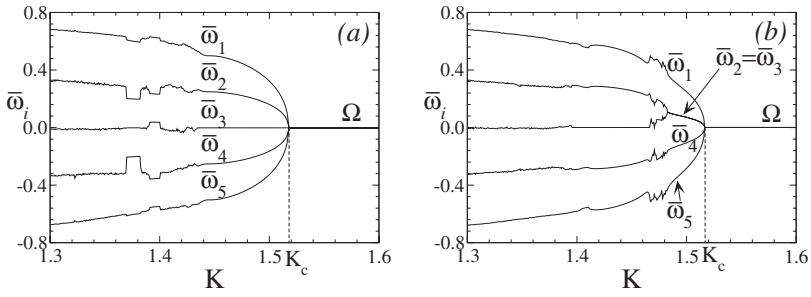


Fig. 5. Frequency-splitting bifurcation diagrams for the Kuramoto model from (1) with $N = 5$ phase oscillators, with a non-uniform distribution of the natural frequencies ω_i: (**a**) $\{-1.0022, -0.50446, 0.00106, 0.49869, 0.99804\}$, and (**b**) $\{-1.00064, -0.50017, 0.00347, 0.50170, 0.99852\}$. $\overline{\omega}_i = \langle \dot{\psi}_i \rangle$, where $\langle \cdot \rangle$ denotes averaging over time. Ω is the mean frequency

observe, with a decrease of the coupling strength K, the common average frequency Ω instantly splits at $K = K_c$ into five and four clusters, respectively. The 1:1:1:1 resonance [Fig. 5(a)] and the 1:1:0:1 resonance [Fig. 5(b)] take place for the corresponding four-dimensional torus dynamics of the phase differences. In Fig. 6 four main resonant tongues with winding ratios 1:1:1:1, 1:0:1:1, 1:1:0:1, and 1:0:1:0 are shown for this case. Each of the tongues hits the frequency-splitting bifurcation curve B_c for a non-empty parameter range. Therefore, for the five-dimensional Kuramoto model, the course of the bifurcation, in the case where more than two phase clusters emerge, is also preserved

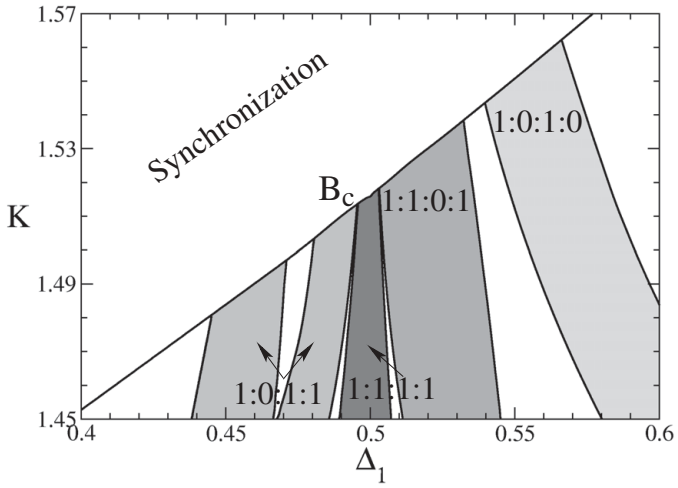


Fig. 6. Two-parameter bifurcation diagram for the Kuramoto model from (1) with N = 5 and $\Delta_i = 0.5, i = \overline{2,4}$. Four main resonant tongues are shown

under small parameter perturbation. In other words, the multiple frequency-splitting bifurcation is of co-dimension 1, as in the case with $N = 3$ described above.

We expect, that this low-dimensional mechanism also controls the frequency splitting bifurcation in the general N-dimensional Kuramoto model from (1). In this case, the dynamics is reduced to an $(N-1)$-dimensional torus flow of the phase differences $\varphi_i = \psi_{i+1} - \psi_i$, which is an $(N-1)$-dimensional analog of (9). For $K > K_c$, synchronization is caused, generally, by a unique sink O. When K decreases to K_c, this sink and, thus, also synchronization vanishes due to an inverse saddle-node bifurcation. For a range of the coupling parameter values greater than K_c, some of the phase differences φ_i already display circular rotations on the $(N-1)$-dimensional torus (like for the Cherry flow in dimension two). Let us characterize these rotations by the $(N-1)$-dimensional analog of the winding ratio, $r_1 : \cdots : r_{N-1}$, which indicates how many phase differences φ_i rotate, and with which ratio with respect to each other. When the steady state O disappears at $K = K_c$, the stable torus rotations arise, in this way causing a desynchronization. As in the three-dimensional case, we expect the number of phase clusters emerging in the N-dimensional frequency-splitting bifurcation to be equal to the number of non-zero coordinates of the corresponding $(N-1)$-dimensional winding ratio $+1$. This mechanism is robust with respect to small parameter perturbations. We conclude that in the Kuramoto model the frequency-splitting bifurcation, in the case where more than two phase clusters emerge, is also of co-dimension 1.

3 Symmetric Kuramoto Model

3.1 Invariant Manifold and Reduction to the Winfree Model

We consider the N-dimensional Kuramoto model of the form (1):

$$\dot{\psi}_i = \omega_i + \frac{K}{N}\sum_{j=1}^{N}\sin(\psi_j - \psi_i)\,, \qquad i = 1, \ldots, N\,. \tag{11}$$

Suppose that the natural frequencies ω_i are symmetrically allocated around the mean frequency Ω. There are two variants for that depending on whether N, the number of oscillators in the ensemble, is even or odd. Let us introduce $N_0 = [N/2]$, where $[\cdot]$ is the integer part of the number. In the even case $N = 2N_0$ and in the odd case $N = 2N_0+1$. By subtracting the mean frequency Ω (by using the variable change $\psi_i \mapsto \psi_i + \Omega t$) the natural frequencies ω_i becomes symmetric with respect to zero as shown in Fig. 7.

The Kuramoto model, with the symmetric natural frequency distribution as in Fig. 7, has an N_0-dimensional invariant manifold M:

$$M = \{\psi_{i+1} - \psi_i = \psi_{-i} - \psi_{-i+1}, \quad i = 1, \ldots, N_0\}\,. \tag{12}$$

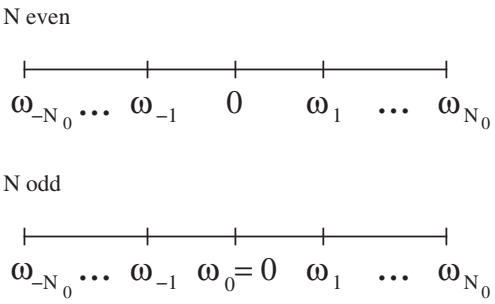


Fig. 7. Schematic representation of the distribution of natural frequencies for the Kuramoto model with even and odd number of oscillators N

The manifold M corresponds to identical deviations of ith and $-i$th phases ψ_i and ψ_{-i} from zero, for any $i = 1, \ldots, N_0$. The average phase Ψ is equal to $(\psi_1 + \psi_{-1})/2$ in the even case, and is equal to $\psi_0 = const$ in the odd case. Hence, Ψ is a constant, and below we can put $\Psi = 0$. Therefore, for the points in the manifold M, the average phase Ψ of the complex order parameter $r(t)\exp(i\Psi)$ can be assumed to equal zero.

The invariance of the manifold M means that any trajectory which is initiated in the manifold will never leave it. The manifold itself is a hyperplane which can be represented as N_0-dimensional torus embedded in the N-dimensional phase space of the Kuramoto model.

The N-symmetric Kuramoto model, being restricted to the manifold M, takes the Winfree form (2):

$$\dot{\psi}_i = \omega_i - Kr\sin\psi_i, \qquad i = 1, \ldots, N_0\ . \tag{13}$$

Here, r is the amplitude of the order parameter (4), which corresponds to the extent of in-phase synchronization:

$$r = \frac{1}{N}\sum_{j=-N_0}^{N_0}\cos\psi_j\ . \tag{14}$$

Therefore, the dynamics inside the manifold M is given by the following N_0-dimensional equations:
$N = 2N_0$:

$$\dot{\psi}_i = \omega_i - \frac{K}{N_0}\sin(\psi_i)\sum_{j=1}^{N_0}\cos\psi_j, \qquad i = 1, \ldots, N_0\ . \tag{15}$$

$N = 2N_0 + 1$:

$$\dot{\psi}_i = \omega_i - \frac{K}{2N_0+1}\sin(\psi_i)\left(2\sum_{j=1}^{N_0}\cos\psi_j + 1\right), \qquad i = 1, \ldots, N_0\ . \tag{16}$$

The reduced (15) and (16) are of Winfree form (2). The natural frequencies ω_i in (15) and (16) are positive (more precisely, not negative). They belong to the right half of the distribution of the natural frequencies of the original N-dimensional Kuramoto model (1) (see Fig. 4). Below let us assume that $\omega_{i+1} \geq \omega_i$, $i = \overline{1, N_0}$.

3.2 Stability of the Invariant Manifold

The existence of the invariant manifold M itself does not guarantee that the dynamics of the Kuramoto model will be restricted to it. For this, the manifold must be transversely stable, i.e. stable with respect to perturbations in directions out of the manifold. The transverse stability of M is controlled by the transverse Lyapunov exponents, i.e. those which correspond to eigenvectors of the Jacobian matrix of the symmetric N-dimensional Kuramoto model which are transverse to the manifold. Note that there are $N - N_0 - 1$ transverse Lyapunov exponents in each point of the manifold M.

The dynamics in the manifold is given by (15) and (16). It is transversely stable if all transverse Lyapunov exponents are negative. The manifold dynamics loses its transverse stability when the maximal transverse Lyapunov exponent becomes positive.

By evaluating the maximal transverse Lyapunov exponent along a typical trajectory of the reduced (15) and (16), parameter regions with transverse stability of M can be obtained. Our numerical simulations show that there exists $K_{sb} < K_c$ such that the dynamics in M is transversely stable for all $K > K_{sb}$. However, as shown in Sect. 3.4 for the case $N = 4$, transversely stable state in the manifold coexists for a range of parameter K values with another chaotic attractor outside the manifold. It would be interesting to verify whether the multistability can take place for larger N.

Since $K_c > K_{sb}$, the synchronized behavior, which is characterized by the common average frequency Ω, is always restricted to the manifold M. With a decrease of K beyond K_c, the synchronization is destroyed in a frequency-splitting bifurcation. The common frequency Ω splits into two or more different frequencies. So for $K_{sb} < K_c$, the bifurcation takes place inside the manifold M. Therefore, the desynchronization transition in the symmetric Kuramoto model is actually determined by the reduced Winfree model given by (15) and (16).

3.3 Desynchronization Transition: From Saddle-Node to Symmetry-Breaking

By a desynchronization transition in the Kuramoto model, we mean a sequence of bifurcations occurring between complete synchronization and complete desynchronization. The latter means that oscillators with different natural frequencies have different average frequencies. After the occurrence of complete desynchronization, other bifurcations may take place. Finally, as

$K \to 0$, the dynamics shows up as an uncoupled behavior acting on the $(N-1)$-dimensional torus.

The desynchronization transition starts at $K = K_c$ with the first frequency-splitting bifurcation. It is caused by the unique stable equilibrium O, which vanishes in a saddle-node bifurcation. The equilibrium belongs to the manifold M, which allows us to find the critical bifurcation parameter analytically. Indeed, any equilibrium of the reduced Kuramoto model (15) and (16), satisfies the equations

$$\omega_i - Kr\sin\psi_i = 0 \tag{17}$$

where

$$r = \begin{cases} \dfrac{1}{N_0}\displaystyle\sum_{j=1}^{N_0}\cos\psi_j, & N = 2N_0 \\ \dfrac{1}{2N_0+1}\left(2\displaystyle\sum_{j=1}^{N_0}\cos\psi_j + 1\right), & N = 2N_0+1\ . \end{cases}$$

As it follows from (17), the equilibria are located on a one-dimensional manifold $\sin\psi_1 = \frac{\omega 1}{\omega_2}\sin\psi_2 = \cdots = \frac{\omega_1}{\omega_{N_0}}\sin\psi_{N_0}$. We denote $x \stackrel{def}{=} \sin\psi_1$. Then x is a root of a scalar algebraic equation

$$x \cdot r(x) = \frac{\omega_1}{K} \tag{18}$$

where

$$r(x) = \begin{cases} \dfrac{1}{N_0}\displaystyle\sum_{j=1}^{N_0}\sqrt{1-\left(\frac{\omega_j}{\omega_1}x\right)^2}, & N = 2N_0 \\ \dfrac{1}{2N_0+1}\left(2\displaystyle\sum_{j=1}^{N_0}\sqrt{1-\left(\frac{\omega_j}{\omega_1}x\right)^2} + 1\right), & N = 2N_0+1\ . \end{cases} \tag{19}$$

To find the bifurcation value $K = K_c$ of the saddle-node bifurcation, we take into account that the stable equilibrium O is given by the smallest positive root of (18). The root disappears (by colliding with another root) when the left-hand part function in (18) reaches the maximum value, i.e., at $x = x_c$, where x_c satisfies

$$x_c r'(x_c) + r(x_c) = 0$$

(r' denotes the derivative of r). Finally, we conclude from (18) that the sadle-node bifurcation occurs at $K = K_c$, where

$$K_c = \frac{\omega_1}{x_c r(x_c)}\ . \tag{20}$$

Formula (20) provides a simple analytical expression for the beginning of the desynchronization transition in the symmetric Kuramoto model. The result

of the bifurcation at $K = K_c$ is the appearance of a stable periodic orbit γ which belongs to the invariant manifold M.

With a further decrease of the coupling strength K, the stable periodic orbit γ undergoes a number of bifurcations, finally producing a chaotic behavior inside the manifold M. The chaotic dynamics in the manifold M first may be transversely stable up to a symmetry-breaking bifurcation value K_{sb}. The system's dynamics changes from a low-dimensional chaotic attractor in the manifold to a high-dimensional hyperchaotic attractor in the whole $(N-1)$-dimensional phase space. As our numerics show, these two attractors, one in the manifold and another outside the manifold, can co-exist for a range of the coupling parameter prior to K_{sb}. Then, the phenomenon of multistability occurs for the Kuramoto model: depending on the initial conditions one of the two stable, very different limiting states is realized. The high-dimensional hyperchaotic behaviour is characterized by the number of positive Lyapunov exponents. We found numerically, that there are two and four positive Lyapunov exponents for $N = 7$ and $N = 10$, respectively.

Below, we illustrate properties of the symmetric Kuramoto model in the cases $N = 4$ and $N = 7$.

3.4 Chaotic Attractor in the Kuramoto Model: $N = 4$

An example of the desynchronization transition for the four-dimensional Kuramoto model is shown in Fig. 8. The corresponding Lyapunov exponents are plotted in Fig. 9. The natural frequencies ω_i are chosen to be uniformly distributed in the interval $[-0.15, 0.15]$.

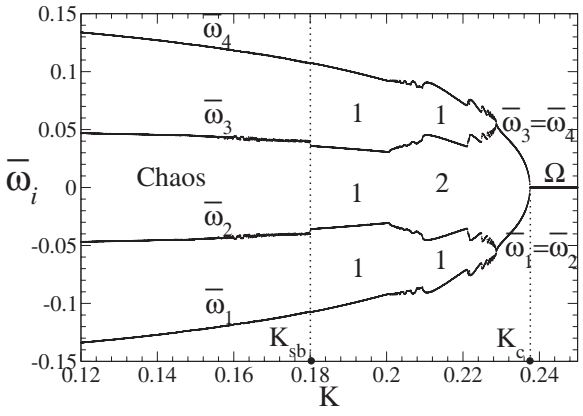


Fig. 8. Frequency-splitting bifurcation diagram for the Kuramoto model from (1) with $N = 4$ phase oscillators, with natural frequencies ω_i uniformly distributed in the interval $[-0.15; 0.15]$. $\overline{\omega}_i = \langle \dot{\psi}_i \rangle$, where $\langle \cdot \rangle$ denotes averaging over time. Ω is the mean of the eigenfrequencies

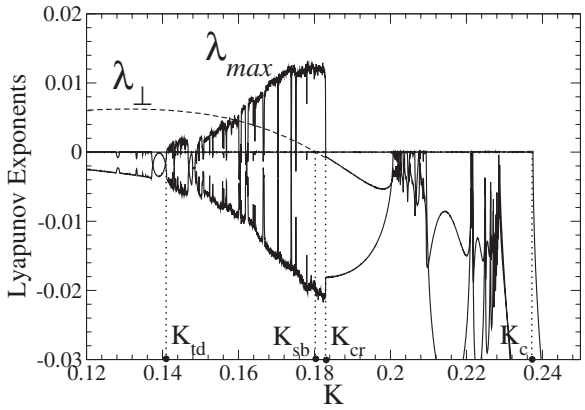


Fig. 9. Lyapunov exponents of the Kuramoto model from (1) with $N = 4$ phase oscillators, with natural frequencies ω_i uniformly distributed in the interval $[-0.15; 0.15]$

It turns out that the maximal Lyapunov exponent is positive for a rather long coupling parameter interval, which indicates the appearance of a chaotic attractor. The chaotic attractor is depicted in Fig. 10. It is born at $K = K_{td} = 0.141\ldots$ due to a torus-destruction scenario [23]. With an increase of

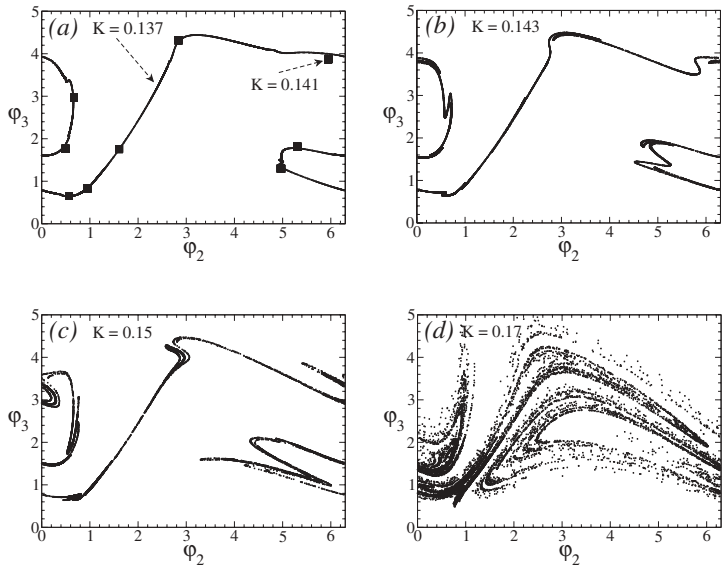


Fig. 10. Return map at $\varphi_1 = 0$ for the Kuramoto model from (1) with $N = 4$ phase oscillators, with a uniform distribution of the natural frequencies ω_i in the interval $[-0.15; 0.15]$

the coupling strength K a fractal structure of the chaotic attractor emerges. Finally, it vanishes at $K = K_{cr} = 0.1815\ldots$ in a boundary crisis.

Before the appearance of the chaotic attractor, i.e. for $K < K_{td}$, the dynamics of the system is periodic or quasiperiodic and may occupy the whole three-dimensional phase space, so that it is not restricted to the invariant manifold. After the disappearance of the attractor, i.e. for $K > K_{cr}$, the dynamics of the system is attracted by the two-dimensional invariant manifold of the form (12)

$$M = \{\psi_2 - \psi_1 = \psi_4 - \psi_3\}$$

and, hence, cannot be chaotic and turns out to be stationary, periodic or quasiperiodic. The manifold dynamics is transversely stable for all $K > K_{sb}$, where $K_{sb} = 0.180\ldots$ is a parameter value of the symmetry-breaking bifurcation of M (i.e. of the transverse destabilization of the in-manifold dynamics).

Note that $K_{sb} < K_{cr}$ which implies multistability in the Kuramoto model. Indeed, in the parameter interval between K_{sb} and K_{cr}, two attractors coexist. One is a stable periodic orbit inside the two-dimensional invariant manifold M, the other is the chaotic attractor in the whole three-dimensional phase space (see Figs. 9 and 10d).

The two-dimensional dynamics inside the manifold is given by (15) with $N_0 = 2$. In the example considered, $\omega_1 = 0.05$ and $\omega_2 = 0.15$. At $K = K_c = 0.24\ldots$ a stable node O is born in a saddle-node bifurcation. It belongs to the manifold and gives rise to the appearance of the Cherry flow for $K > K_c$. The Cherry flow has two fixed points, a stable node, a saddle, and an unstable periodic orbit of winding ratio 1:0.

To trace the desynchronization process in the original four-dimensional Kuramoto model, we decrease the coupling strength back from the Cherry flow to the three-dimensional chaotic attractor (see. Figs. 8 and 10). First, at $K = K_c$, the frequency-splitting bifurcation occurs. A stable periodic orbit with winding ratio 1:0:1 arises and provides two phase clusters: $\bar{\omega}_1 = \bar{\omega}_2$ and $\bar{\omega}_3 = \bar{\omega}_4$. At the next bifurcation moment, at $K = 0.228..$, the phases in both clusters split from each other. Beyond the splitting, the parameter point crosses infinitely many higher resonant tongues.

At $K = 0.2\ldots$ the system approaches the main resonant tongue 1:1:1. The stable periodic behavior inside it is still transversely stable and hence, as before, the system's dynamics takes place in the manifold M. With decreasing K the resonant behavior within the manifold loses its transverse stability, which causes a symmetry-breaking bifurcation at $K = K_{sb}$, as shown by the graph of the transverse Lyapunov exponent $\lambda_\perp$ in Fig. 9. After the symmetry-breaking, the dynamics of the system is no longer restricted to the manifold and is captured by the chaotic attractor.

With a further decrease of the coupling strength beyond K_{td}, the behavior is mostly quasiperiodic on a two-dimensional torus which lies outside the manifold M. The strength of the attraction of the torus decreases with decreasing K. Finally, as K tends to zero, one gets an uncoupled behavior on

the three-dimensional torus. Its winding ratio is given by the differences of the natural frequencies of the Kuramoto model. In the example considered, we again obtain the resonance 1:1:1.

3.5 Hyperchaos: $N = 7$

In Figs. 11 and 12, an example of a desynchronization transition and the graphs of the corresponding Lyapunov exponents are shown for the Kuramoto model of dimension $N = 7$. Again, we examine the case with natural frequencies equally distributed, now in the interval $[-1\ ;\ 1]$.

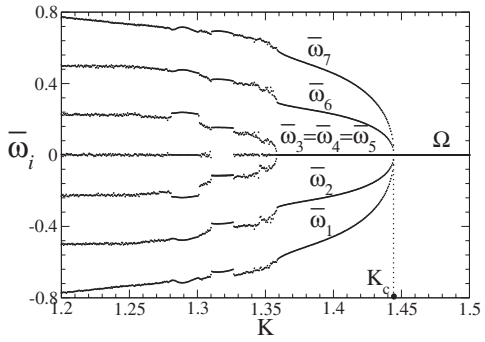


Fig. 11. Frequency-splitting bifurcation diagram for the Kuramoto model from (1) with $N = 7$ phase oscillators, with natural frequencies ω_i uniformly distributed in the interval $[-1.0; 1.0]$. $\overline{\omega}_i = \langle \dot{\psi}_i \rangle$, where $\langle \cdot \rangle$ denotes averaging over time. Ω is the mean frequency

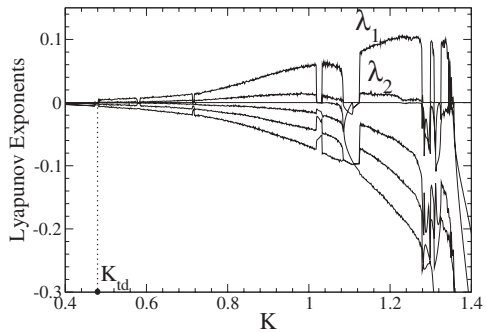


Fig. 12. Lyapunov exponents for the Kuramoto model from (1) with $N = 7$ phase oscillators, with natural frequencies ω_i uniformly distributed in the interval $[-1.0; 1.0]$

The symmetric invariant manifold M is three-dimensional. The dynamics within this manifold is given by (16) with $N_0 = 3$ and natural frequencies $\omega_1 = 1/3$, $\omega_2 = 2/3$, and $\omega_3 = 1$. It can be regular or chaotic. For $K = K_{sb} = 1.28\ldots$, a hyperchaotic behavior in the whole six-dimensional phase space emerges. Indeed, two maximal Lyapunov exponents λ_1 and λ_2 are positive for a sufficiently long parameter range (see Fig. 12). They vanish only at $K = K_{td} \sim 0.477$.

4 Conclusion

In the present chapter, we considered the Kuramoto model from the standpoint of nonlinear dynamical systems. Our approach revealed a sequence of bifurcations, which, in particular, give rise to the appearance of desynchronized states, chaos, hyperchaos, and multistability. The desynchronization transition starts with the saddle-node bifurcation of the unique phase-locked state. After the bifurcation phase clusters emerge, each rotating with its own frequency. With a further decrease of the coupling strength, the bifurcation scenario, which we present here, finally leads to complete phase desynchronization, where all phases rotate with different frequencies. On this route to desynchronization, with a decrease of the coupling strength, first, the dynamics is regular, namely periodic or quasiperiodic. This regularity ceases with the transition to chaos. The chaotic regime is intersected by windows of periodicity.

The Kuramoto model displays high-dimensional hyperchaotic behavior, which is present within a wide parameter range. The complexity of the hyperchaos can be characterized by the number of positive Lyapunov exponents. This number grows with the dimension N of the Kuramoto model, which indicates that the hyperchaos becomes more and more developed with increasing system size N. We observed, that there are two positive Lyapunov exponents in the case with $N = 7$; four for $N = 10$; and nine for $N = 20$. In the case $N = 4$, we have also observed multistability of chaotic and regular dynamics.

The richness and complexity of the nonlinear dynamics of the Kuramoto model is interesting and highly significant from the standpoint of both theory and applications. Our results demonstrate how fruitful the nonlinear dynamics approach may contribute to a better understanding of the behavior of the Kuramoto model.

References

1. A.T. Winfree, J. Theoret. Biol. **16**, 15 (1967). 287
2. A.T. Winfree, *The Geometry of Biological Times* (Springer, New-York, 1980). 285, 287
3. Y. Kuramoto, *Chemical Oscillations, Waves and Turbulence* (Springer-Verlag, Berlin, 1984). 285, 286

4. G.B. Ermentrout, Physica D **41**, 219 (1990). 285
5. H. Daido, Int. J. Bifurcation Chaos **7**, 807 (1997) and references therein. 285
6. L. Ren and B. Ermentrout, Physica D **143**, 56 (2000). 285
7. J.T. Ariaratnam and S.H. Strogatz, Phys. Rev. Lett. **86**, 4278 (2001). 285, 287
8. J.A. Acebron, A. Perales, and R. Spigler, Phys. Rev. E **64**, 016218 (2001). 285
9. E. Ott, P. So, E. Barreto, and T. Antonsen, Physica D **173**, 29 (2002). 285
10. K. Wiesenfeld, P. Colet, and S.H. Strogatz, Phys. Rev. Lett. **76**, 404 (1996); Phys. Rev. E **57**, 1563 (1998). 285
11. G. Kozyreff, A.G. Vladimirov, and P.Mandel, Phys. Rev. Lett. **85**, 3809 (2000). 285
12. W. Wang, I.Z. Kiss, and J.L. Hudson, Chaos **10**, 248 (2000); Phys. Rev. Lett. **86**, 4954 (2001). 285
13. C.S. Peskin, *Mathematical Aspects of Heart Physiology* (Courant Institute of Mathematical Sciences, New-York, 1975). 285
14. J. Buck, Q. Rev. Biol. **63**, 265 (1988). 285
15. P.A. Tass, Phys. Rev. E **56**, 2043 (1997). 288
16. P.A. Tass, *Phase Resetting in Medicine and Biology* (Springer, Berlin, 1999). 285, 288, 290
17. P.A. Tass, Europhys. Lett. **53**, 15 (2001); Europhys. Lett. **55**, 171 (2001); Biol. Cybern. **57**, 164 (2002); Europhys. Lett. **53**, 15 (2001); Biol. Cybern. **87**, 102 (2002); Phys. Rev. E **66**, 036226 (2002); Biol. Cybern. **89**, 81 (2003). 285, 289, 290
18. G.B. Ermentrout and N. Kopell, SIAM J. Math. Anal. **15**, 215 (1984). 286
19. S.H. Strogatz, Physica D **143**, 1 (2000). 286, 287
20. A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization, A Universal Concept in Nonlinear Sciences* (Cambridge University Press, 2001); D.Topaj, and A.Pikovsky, Physica D **170**, 118 (2002). 286
21. M. Zhan, Z. Zheng, G.Hu, and X. Peng, Phys. Rev. E **62**, 3552 (2000); Z. Zheng, G. Hu, and B. Hu, Phys. Rev. E **62**, 7501 (2000) 286
22. Yu. Maistrenko, O. Popovych, O. Burylko, and P.A. Tass, Phys. Rev. Lett. (2004) (to appear). 286
23. V.S. Afraimovich, L.P. Shilnikov, Am.Math.Soc.Trans **149**, 176 (1991); V.I. Arnold, V.S. Afraimovich, Yu.S. Ilyashenko, L.P. Shilnikov, Encyclopedia of Mathematical Sciences, Dynamical Systems V (Springer-Verlag, Berlin, 1994). 286, 302
24. D. Hansel, G. Mato and C. Meunier, Phys. Rev. E **48**, 3470 (1993). 289
25. H. Kori and Y. Kuramoto, Phys. Rev. E **63**, 046214 (2001). 289
26. H.G. Schuster and P. Wagner, Prog. Theor. Phys. **81**, 939 (1989). 289
27. S. Kim, S.H. Park, and C.S. Ryu, Phys. Rev. Lett. **79**, 2911 (1997); M.K.S. Yeung and S.H. Strogatz, Phys. Rev. Lett. **82**, 648 (1999); M.G. Rosenblum, A.S. Pikovsky, Phys. Rev. Lett. **92**, 114102 (2004). 289
28. A. Nini et al., J. Neurophysiol. **74**, 1800 (1995). 289
29. A.L. Benabid et al., The Lancet **337**, 403 (1991); S. Blond et al., J. Neurosurg. **77**, 62 (1992). 290
30. P.A. Tass, J. Biol. Phys. **22**, 27 (1996); J. Biol. Phys. **22**, 125 (1996). 290
31. T.M. Cherry, Proc. London Math. Soc. **44**, 175 (1938). 294
32. C. Boyd, Ergod. Th. Dyn. Syst. **5**, 27 (1985). 294, 295
33. J.J.P. Veerman, Nonlinearity **2**, 419 (1989). 294, 295
34. C. Baesens, J. Guckenheimer, S. Kim, R.S. MacKay, Physica D **49**, 387 (1991). 294
35. M.A. Zaks, Physica D **149**, 237 (2001). 294, 295

Spatially Extended Monotone Mappings

R. Coutinho[1] and B. Fernandez[2]

[1] Departamento de Matemática, Instituto Superior Técnico, Av. Rovisco Pais 1096, Lisboa Codex, Portugal
Ricardo.Coutino@math.ist.utl.pt

[2] Centre de Physique Théorique, CNRS & Universités de Marseille et de Toulon, Luminy Case 907, 13288 Marseille Cedex 09, France
bastien@cpt.univ-mrs.fr

1 Introduction

This chapter deals with the study of travelling waves in discrete time spatially extended systems with monotone dynamics. Such systems appear for instance in alloy solidification, in population dynamics and in solid-state physics. Special emphasis is made on the existence of travelling waves, on the uniqueness of their velocity and on their relevance for the description of propagation phenomena in such systems.

The first section deals with interfaces between two stable homogeneous phases and their propagation in the form of fronts. The analysis applies to systems of bistable one-dimensional maps coupled via the convolution with an arbitrary distribution function [6]. This analysis completes a previous work on piecewise affine bistable CML [3, 4] and its extension to systems of piecewise affine one-dimensional maps coupled via convolution [5].

The second section deals with travelling waves in monotonous extended systems driven by spatially periodic forces. These systems are inspired by discrete time analogues of the dissipative dynamics of Frenkel-Kontorova models (see the chapters by Floría, Baesens and Gómez-Gardeñes and by Baesens for such dynamics in continuous time). For such nonlinear systems, a dispersion relation is obtained and the existence of travelling waves with arbitrary wave and corresponding frequency is shown [7].

In spite of similarities with other works in the literature (see e.g. [12]) the methods and, particularly, the formalism developed in the papers [3, 4, 5, 6, 7] are quite distinct and original. They encompass in a unified framework, systems with continuous couplings and systems with discrete couplings. In particular, changes in the dynamics of travelling waves (e.g. changes in shape and in velocity) are described when the coupling continuously changes from a discrete to a continuous one.

R. Coutinho and B. Fernandez: *Spatially Extended Monotone Mappings*, Lect. Notes Phys. **671**, 265–284 (2005)
www.springerlink.com

2 Bistable Extended Maps

2.1 From Bistable CML to General Bistable Extended Mappings

As a starting point, we consider the basic model of CML

$$u_s^{t+1} = (1-\varepsilon)f(u_s^t) + \frac{\varepsilon}{2}\left(f(u_{s-1}^t) + f(u_{s+1}^t)\right) \tag{1}$$

where the real number $u_s^t \in [0,1]$ represents the element of a lattice configuration at discrete time $t \in \mathbb{Z}$ and discrete space $s \in \mathbb{Z}$, and where $\varepsilon \in [0,1)$ is the coupling strength. The map f is a **bistable map** from $[0,1]$ into itself. A bistable map is a continuous increasing mapping from $[0,1]$ into itself with exactly 3 fixed points, namely the points 0, c and 1. The points 0 and 1 are attracting and c is unstable, see Fig. 1.

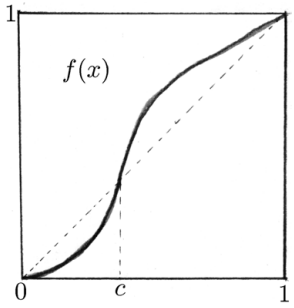


Fig. 1. A bistable map f

The evolution of a configuration under a bistable CML can be viewed as a local force which impels convergence to some steady state (a stable fixed point of f, either 0 or 1), followed by a diffusion process which takes the form of the following linear operator

$$(\mathrm{L}u)_s = (1-\varepsilon)u_s + \frac{\varepsilon}{2}(u_{s+1} + u_{s-1})\ . \tag{2}$$

In particular, if at time t the configuration satisfies $u_s^t > c$ for all $s \in \mathbb{Z}$, then the evolution forces the configuration to converge uniformly to the fixed point $u_s = 1$ for all $s \in \mathbb{Z}$. On the opposite, if $u_s^t < c$ for all $s \in \mathbb{Z}$, then the evolution forces the configuration to converge uniformly to the fixed point $u_s = 0$ for all $s \in \mathbb{Z}$. Hence these fixed points represent stable phases of the dynamics.

The goal of this first section is to investigate the competition between these two phases, namely the evolution of configurations which at some time satisfy $u_s^t < c$ for sufficiently large negative s and $u_s^t > c$ for sufficiently large positive s. Typically, the resulting motion is the invasion of one phase into the other one, an invasion ruled by special solutions of (1), namely the fronts.

A front of rational velocity $\frac{p}{q}$ for the CML (1) is a configuration which satisfies this evolution equation and the relations

$$\forall s,t : u_s^{t+q} = u_{s-p}^t, \quad \lim_{s\to-\infty} u_s^t = 0 \quad \text{and} \quad \lim_{s\to+\infty} u_s^t = 1 \,.$$

Before starting the investigation of such fronts, one may wonder about the existence of fronts with irrational velocity in CML. A simple way to take into account both rational and irrational velocities in a unique formalism is to define a **front** of velocity v for the CML (1) as a configuration which satisfies

$$u_s^t = \phi\,(s - tv)$$

where $\phi : \mathbb{R} \to [0,1]$ is a real function, called the **front shape**, which satisfies

$$\lim_{x\to-\infty} \phi\,(x) = 0 \quad \text{and} \quad \lim_{x\to+\infty} \phi\,(x) = 1 \,.$$

In other words the introduction of functions of real variable in CML allows to study fronts with irrational velocities. However, using also this formalism in the description of fronts with rational velocities has proved to be useful for the global comprehension of fronts dynamics.

Early studies in this direction showed the existence of such fronts (of rational or irrational velocity depending on parameters) for (1) when f is a discontinuous piecewise affine map with constant slope [3]. The existence of fronts has also been shown in a CML with continuous piecewise affine bistable local map and unidirectional coupling [1]. The proof in [3] is achieved by using an explicit construction of the front shape for this CML. In order to show the existence of such fronts when the local map f is an arbitrary (continuous) bistable map, it is useful to fully generalise the model under consideration.

Actually, one may not restrict oneself to CML defined by (1) but also consider more general CML whose iterations involve larger, even unbounded, neighbourhoods. In addition, a reversal symmetric coupling is not required and this assumption can be dropped. That is to say, one may consider the following coupling operator

$$(\mathrm{L}u)_s = \sum_{n\in\mathbb{Z}} \ell_n u_{s-n}$$

where the coefficients ℓ_n are nonnegative real numbers such that

$$\sum_{n\in\mathbb{Z}} \ell_n = 1 \,.$$

Note that the coupling operator in (2) is recovered by choosing $\ell_{-1} = \ell_1 = \frac{\varepsilon}{2}$, $\ell_0 = 1 - \varepsilon$ and $\ell_n = 0$ for $n \notin \{-1, 0, 1\}$.

Since the front shape is defined as a function of real variable, it is natural to extend the action of the dynamics to the functions of real variable. That is to say instead of the CML (1), we consider

$$u^{t+1}(x) = \sum_{n\in\mathbb{Z}} \ell_n f(u^t(x-n)) .$$

where each u^t is a function from $\mathbb{R}$ to $[0,1]$.

Of course, the dynamics of lattice configurations is recovered by considering the invariant set of functions which are constant on every interval $[s, s+1)$ where $s \in \mathbb{Z}$. But such an extension provides an appropriate framework to the front shape dynamics.

At this stage, an additional extension appears immediately. One may consider diffusive linear operators of the form

$$(\mathrm{L}u)(x) = \sum_{n\in\mathbb{Z}} \ell_n u(x-r_n)$$

where the coefficients ℓ_n are nonnegative real numbers such that $\sum_{n\in\mathbb{Z}} \ell_n = 1$ and r_n are arbitrary real numbers (not only integers). An alternative way of writing this operator is by using the **convolution** with a distribution function h. Recall that such a convolution is defined by the Lebesgue-Stieltjes integral

$$(\mathrm{L}u)(x) = h * u(x) := \int_{\mathbb{R}} u(x-y)dh(y) . \tag{3}$$

That is to say, one may consider diffusive linear operator of the form $Lu = h*u$. In the previous case, the coupling operator can be written $h * u$ with the distribution function h being given by $h(x) = \sum_{n\in\mathbb{Z}} \ell_n H(x-r_n)$ where H is the Heaviside function

$$H(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ 1 & \text{if} \quad 0 \le x \end{cases}$$

Convolutions are not limited to discrete distribution functions and, as a final extension, one may consider $Lu = h * u$ where h is an arbitrary distribution function, that is to say, any increasing function with the following limits $\lim_{x\to-\infty} h(x) = 0$ and $\lim_{x\to+\infty} h(x) = 1$.

Therefore instead of only considering the dynamics of fronts in bistable CML, we consider the dynamics of fronts in bistable (spatially) extended maps whose iterations write

$$u^{t+1} = Fu^t = h * f \circ u^t \tag{4}$$

where h is an arbitrary distribution function and f an arbitrary bistable map. (A further extension will be considered in Sect. 2.4.)

In order to have a well-defined convolution operator (3) for arbitrary distribution function h, the functions u need be Borel measurable. Accordingly we consider the dynamics (4) in the set $\mathcal{B}$ of Borel-measurable functions defined on $\mathbb{R}$ with values in $[0,1]$.

It is noteworthy that the present formalism collects in a unique framework, CML and classical models with continuous diffusive couplings. On one hand the basic model of CML is recovered for

$$h(x) = (1-\varepsilon)H(x) + \frac{\varepsilon}{2}(H(x+1) + H(x-1)) \ .$$

On the other hand by choosing h to be the absolutely continuous distribution function with heat kernel

$$h(x) = \int_{-\infty}^{x} e^{-\pi y^2} dy, \quad x \in \mathbb{R}$$

the map $Fu(x) = (h * f \circ u)(x) = \int_{\mathbb{R}} f \circ u(x-y) e^{-\pi y^2} dy$ gives an integral formulation of a reaction-diffusion process in discrete time. Indeed, we then have $Fu(x) = w_u(x, 1/4\pi)$, where $w_u(x,t)$ is the solution of the initial value problem $\frac{\partial w_u}{\partial t} = \frac{\partial^2 w_u}{\partial x^2}$ and $w_u(x,0) = f \circ u(x)$.

The rest of this section is dedicated to a sketch of the front analysis in bistable extended maps defined by (4). This amounts to prove the existence of a velocity v and of a front shape ϕ which solves the front equation

$$\phi(x - v) = h * f \circ \phi(x) \ .$$

2.2 Basic Properties

Every bistable extended map F defined by (4) has three basic properties. The first property, which is intensively used in the analysis, is **homogeneity**. Homogeneity is expressed by the relation

$$T^v F = F T^v \quad \text{for all } v \in \mathbb{R} \tag{5}$$

where T^v is the **translation** by v, namely the operator acting in $\mathcal{B}$ and defined by $T^v u(x) = u(x - v)$ for all $x \in \mathbb{R}$.

The second important property of F is **continuity** in the sense of pointwise convergence, namely

$$\forall x \in \mathbb{R} \quad \lim_{n\to\infty} u_n(x) = u(x) \quad \Rightarrow \quad \forall x \in \mathbb{R} \quad \lim_{n\to\infty} Fu_n(x) = Fu(x) \ . \tag{6}$$

Finally, the third fundamental property is **monotonicity**, namely

$$u \le v \quad \Rightarrow \quad Fu \le Fv \ . \tag{7}$$

Using these three properties we can deduce other important facts:

(a) Under the action of F, every increasing function is mapped into an increasing function. Using the fact a function is increasing iff it lies above any of its right translation, we have

$$\forall \delta > 0 \quad T^\delta u \le u \quad \Rightarrow \quad T^\delta Fu \le Fu \ .$$

(b) The map F commutes with the **projection** P_ℓ on left continuous functions. Indeed this projection is defined by $P_\ell u(x) = \lim_{\substack{y\to x \\ y<x}} u(y)$ and for any function in its domain, we have $P_\ell u(x) = \lim_{n\to\infty} u(x - \frac{1}{n})$. Using homogeneity and continuity, we conclude that $F P_\ell u = P_\ell F u$. Similarly, one shows that F commutes with the projection P_r on right continuous functions.

2.3 Results and Concepts

Existence of Fronts

As suggested before, any bistable extended map of the form (4) has fronts for some velocity v. This is formally claimed in the following statement.

Theorem 2.1. *For any distribution function h and any bistable map f, there exists a velocity $v \in \mathbb{R}$ and an increasing function ϕ with the following limits $\lim_{x\to-\infty}\phi(x) = 0$ and $\lim_{x\to+\infty}\phi(x) = 1$ which solves the front equation $T^v\phi = h * f \circ \phi$.*

The proof of this Theorem is sketched in Sect. 2.5. The proof is accomplished by a construction of an increasing function ϕ with the desired properties for a chosen velocity v.

Note that monotonicity of the front shape ϕ is not imposed by the front equation. Indeed in some cases (e.g. in weakly coupled CML or in strongly unidirectionally coupled CML), it may happen that a bistable extended map also has fronts with non monotonous shape.

Moreover, uniqueness of monotonous shape cannot be expected in general. There are examples of bistable CML with several monotone front shapes which cannot be identified by applying translations.

Bistable Regular Maps and the Uniqueness of Front Velocity

In spite of the front shape need not be unique, one may wonder about the uniqueness of the velocity. It turns out that this uniqueness holds provided that the map f is so-called regular, a fairly general situation. Indeed, a bistable map f is said to be **regular** if it is a weak contraction in the neighbourhoods of the stable fixed points (see Fig. 2), i.e. if we have

$$\exists\delta > 0 \quad [x, y \in (0, \delta) \quad \text{or} \quad x, y \in (1-\delta, 1)] \quad \Rightarrow |f(x) - f(y)| \leq |x - y| \ .$$

That a bistable map be regular is a quite mild condition relies on Taylor expansion. Indeed by using Taylor formula, one obtains the following sufficient conditions for a bistable map f to be regular

f is analytic, or
$f \in C^1$ and $f'(0) < 1$ and $f'(1) < 1$, or
$f \in C^2$ and $f''(0) \neq 0$ and $f''(1) \neq 0$, or
$f \in C^3$ and $f'''(0) \neq 0$ and $f'''(1) \neq 0$, and
...

Nevertheless, one can exhibit examples of non regular C^∞ bistable maps.

With regularity provided, Theorem 2.1 can be completed by an assertion on velocity uniqueness.

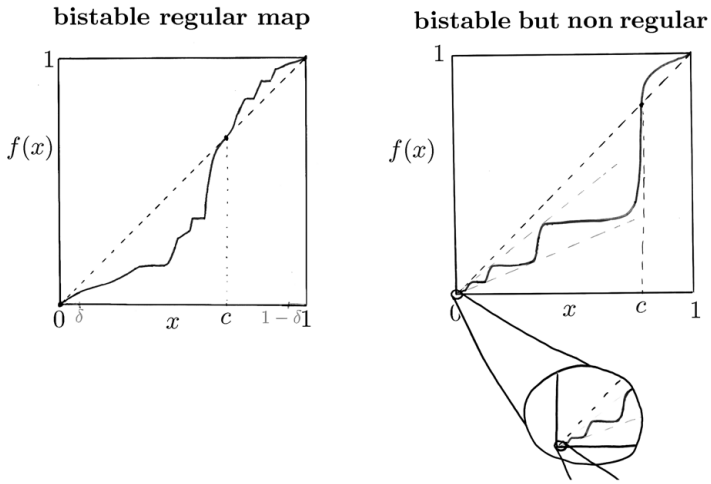


Fig. 2. Bistable regular and non regular local maps

Theorem 2.2. *For any distribution function h and any* regular *bistable map f, there exists a* unique *velocity $v \in \mathbb{R}$ and an increasing function ϕ with the following limits* $\lim_{x\to-\infty}\phi(x) = 0$ *and* $\lim_{x\to+\infty}\phi(x) = 1$ *which solves the front equation* $T^v\phi = h * f \circ \phi$.

Hausdorff Distance of Increasing Functions

Uniqueness of the front velocity in extended systems with regular bistable maps naturally addresses the question of the dependence of this velocity on the local map and on the coupling operator, i.e. the dependence of $v(f, h)$ on f and on h.

The dependence under consideration here is continuity with changes in f and in h. In order to address this problem, adequate notions of convergence both for local maps f and for distribution functions h need to be introduced.

As far as local maps are concerned, convergence is understood in the pointwise sense, i.e. $\{f_n\}$ converges to f iff $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in [0, 1]$.

As far as distribution functions are concerned, the convergence is ruled by a distance in the set of (right continuous) increasing functions. The distance between two distribution functions h and h' is given by

$$d(h, h') = \inf\{\varepsilon > 0 \ : \ h(x-\varepsilon) - \varepsilon \le h'(x) \le h(x+\varepsilon) + \varepsilon, \quad \forall x \in \mathbb{R} \ \} \ . \quad (8)$$

For this distance, the ball of radius ε centred at h is the set of functions for which the graph lies in the band of width $2\sqrt{2}\varepsilon$ in the direction of the line $y = -x$ around the graph Gh of h. The graph of h is defined by

$$Gh = \{(x, y) : P_\ell h\,(x) \le y \le P_r h\,(x)\} \ .$$

In fact it is not difficult to show that this distance is the Hausdorff distance restricted to graphs of such functions:

$$d(h,h') = \max\left\{ \sup_{z_1 \in Gh} \left(\inf_{z_2 \in Gh'} \|z_1 - z_2\| \right), \sup_{z_1 \in Gh'} \left(\inf_{z_2 \in Gh} \|z_1 - z_2\| \right) \right\}$$

where the $\mathbb{R}^2$ norm $\|.\|$ is given by $\|(x,y)\| = \max\{|x|, |y|\}$.

By using relation (8), one shows that the convergence with respect to the distance $d(\cdot,\cdot)$ is equivalent to the convergence at all continuity points, the usual convergence of distribution functions [10]. Precisely, we have $\lim_{n\to\infty} d(h_n,h) = 0$ iff $\lim_{n\to\infty} h_n(x) = h(x)$ for all x where h is continuous [6].

The advantage of such a distance is that it allows continuous distribution functions to converge to discontinuous ones and vice-versa. As a particular consequence, changes in front velocity can be analysed when continuously passing from a model with continuous diffusive operator to a model with discrete diffusive operator (and vice-versa).

Continuity of the Front Velocity

The continuous dependence of the front velocity with respect both to the local map and to the distribution function is given by the following statement. Assume that f is regular and let $v(f,h)$ be the unique front velocity of the mapping $Fu = h * f \circ u$.

Theorem 2.3. *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of regular bistable maps which converges pointwise to a bistable regular map f. Let $\{h_n\}_{n\in\mathbb{N}}$ be a sequence of distribution functions and h be a distribution function such that $\lim_{n\to\infty} d(h_n,h) = 0$. Then $\lim_{n\to\infty} v(f_n,h_n) = v(f,h)$.*

In particular, the front velocity varies continuously with any parameter of the local map (provided that the map pointwise depends continuously on its parameter(s)) and with any coupling parameter (provided the distribution function depends continuously on this parameter). For instance, in the CML defined by (1), the front velocity depends continuously on ε.

A special consequence of this result is the existence of fronts with irrational velocity (for appropriate value of ϵ) in any CML (1) for which the front velocity is not constant when ε moves in $[0,1]$.

As suggested before, Theorem 2.3 contains the claim that the front velocity of an extended bistable map with infinite range coupling can be approximated to arbitrary accuracy by the front velocity of an extended bistable map with finite range discrete coupling, and vice-versa.

Interfaces and Reference Centres $J_a(\psi)$

Once the existence of front has been established, the natural question to address is their Lyapunov stability. Lyapunov stability of fronts is an elaborated question which lies beyond the scope of this chapter.

In this section we provide some information about the dynamics of configurations which need not be monotone, nor need to cross once the unstable fixed point c of f. These configurations are called interfaces. An **interface** is a function $u \in \mathcal{B}$ such that there exists $c_- \in (0, c)$, $c_+ \in (c, 1)$ and $j_1 \leq j_2 \in \mathbb{R}$ so that $u(x) \leq c_-$ if $x \leq j_1$ and $u(x) \geq c_+$ if $x \geq j_2$ (see Fig. 3 for an example of an interface crossing several times the point c.)

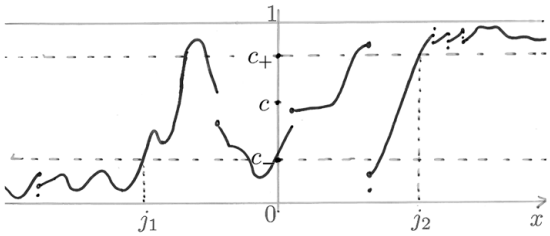


Fig. 3. An interface function u

Interfaces possess the following dynamical properties. If u is an interface, then every iteration $F^t u$ $(t \geq 0)$ is an interface. Moreover, the numbers c_- (resp. c_+) can be chosen arbitrarily near to 0 (resp. 1) provided that t is chosen (accordingly) large enough.

As shown below, the asymptotic dynamical property shared by all interfaces is a unique propagation velocity, the front velocity of course.

In order to compute this velocity, the interface location at each time is measured according to a reference threshold a. We introduce the **reference centre** of an interface u as the smallest point at which the function is not smaller than a, i.e.

$$J_a(u) = \inf \{x \in \mathbb{R} \ : \ u(x) \geq a\} \ .$$

In the case where this quantity is finite, by applying a translation, the function can be centred at 0. Indeed, we have $J_a(T^{-J_a(u)}u) = 0$.

Velocity of Interfaces

According to the previous dynamical property, for any interface and any $a \in (0, 1)$, the quantity $J_a(F^t u)$ is finite for all t sufficiently large. The next statement claims that any interface has asymptotically the front velocity, no matter the initial number of crossing the level c is.

Theorem 2.4. *Let h be a distribution function and let f be a regular bistable map. For every interface u and every $a \in (0,1)$, we have*

$$\lim_{t\to+\infty} \frac{J_a(F^t u)}{t} = v(f,h) \; .$$

Needless to say that the front velocity is an important characteristic of extended bistable map. It plays a similar role to the one played by the rotation number in circle maps.

2.4 Generalisation

The results on front dynamics extend to linear convex combinations of maps of the form $Fu = h * f \circ u$. An interesting application of such an extension resides in lattice dynamical systems as introduced in the chapter by Bunimovich in this volume.

Instead of the map F defined by (4) we now consider the map F defined by

$$Fu = \sum_{k\in\mathbb{N}} a_k h_k * f_k \circ u, \quad u \in \mathcal{B} \; .$$

Here the numbers $a_k \geq 0$ and $\sum_{k\in\mathbb{N}} a_k = 1$. The functions h_k are distributions functions and the maps f_k are continuous increasing maps defined on $[0,1]$ such that there exists $c \in (0,1)$ so that for every $k \in \mathbb{N}$ we have

$$f_k(x) \leq x \quad \text{if } 0 \leq x \leq c \quad \text{and} \quad x \leq f_k(x) \quad \text{if } c \leq x \leq 1 \; .$$

Moreover, we assume that the map

$$f = \sum_{k\in\mathbb{N}} a_k f_k$$

is bistable. Its unstable fixed point is then c.

In addition, we say that the map F is regular if there exists $\delta > 0$ such that for every $k \in \mathbb{N}$ we have

$$|f_k(x) - f_k(y)| \leq |x - y| \quad \text{if } x, y \in (0,\delta) \quad \text{or if } x, y \in (1-\delta, 1) \; .$$

All previous results on existence of fronts (Theorem 2.1), uniqueness of the velocity (Theorem 2.2), continuous dependence of the velocity on the parameters (Theorem 2.3) and existence and uniqueness of the velocity of interfaces (Theorem 4) extend to the present mapping F.

Example. Lattice dynamical system. Let $\varepsilon \in (0,1)$ and f be a regular bistable map such that the map f_0 defined on $[0,1]$ by $f_0(x) = \frac{f(x)-\varepsilon x}{1-\varepsilon}$ is increasing. Let $f_1(x) = x$, $a_0 = 1-\varepsilon$, $a_1 = \varepsilon$, $a_k = 0$ if $k > 1$, $h_0 = H$ and $h_1 = \frac{1}{2}(T^1 H + T^{-1} H)$.

The map

$$Fu(x) = \sum_{k\in\mathbb{N}} a_k h_k * f_k \circ u(x) = f \circ u(x) + \frac{\varepsilon}{2}(u(x-1) - 2u(x) + u(x+1))$$

satisfies the desired properties and is regular.

2.5 Sketch of the Proof of Existence of Fronts

This section presents a brief description of the proof of front existence. The complete proof is given in [6].

The first step consists in introducing **subfronts**, that is to say increasing functions which satisfy the inequality $Fu \leq T^v u$. To be precise, let $\mathcal{I} \subset \mathcal{B}$ be the subset composed of increasing functions, let $v \in \mathbb{R}$ and $c_+ \in (c, 1)$. Consider the set of subfronts of velocity v defined by

$$\mathcal{S}_{v,c_+} = \left\{ \psi \in \mathcal{I} \ : \ F\psi \leq T^v \psi \text{ and } J_{c_+}(\psi) = 0 \right\} .$$

If $\mathcal{S}_{v,c_+}$ is not empty, consider the function

$$\eta_v(x) = \inf_{u \in S_{v,c+}} u(x), \quad x \in \mathbb{R} .$$

It turns out that $\eta_v \in \mathcal{S}_{v,c_+}$ and therefore η_v is a minimal sub-front of velocity v.

In a second step, we consider the maximal sub-fronts velocity

$$\bar{v} = \sup \left\{ v \in \mathbb{R} \ : \ \mathcal{S}_{v,c_+} \neq \emptyset \right\}$$

and we consider the minimal sub-front of maximal velocity, namely $\eta_{\bar{v}}$. (This minimal subfront exists because one shows that $S_{\bar{v},c_+} \neq \emptyset$.) The construction suggests $\eta_{\bar{v}}$ is a good candidate to solving the front equation. However, this is not always the case.

In order to construct a front shape from this function, we start by computing iterates $F^n \eta_{\bar{v}}$. We translate them so that they all be centred at 0 (i.e. $J_{c_+}(T^{-j_n} F^n \eta_{\bar{v}}) = 0$ where $j_n = J_{c_+}(F^n \eta_{\bar{v}})$) and we look for a limit function. That is to say, we consider the sequence $\{T^{-j_n} F^n \eta_{\bar{v}}\}_{n \in \mathbb{N}}$.

Then we prove that $\liminf_{n \to \infty} (j_{n+m} - j_n) = m\bar{v}$. This property is employed together with an arithmetical lemma in order to ensure the existence of a strictly increasing sequence $\{n_k\}$ such that for all m we have

$$\lim_{k \to \infty} (j_{n_k+m} - j_{n_k}) = m\bar{v} .$$

By Helly's Selection Theorem[3] the sequence $\{T^{-j_{n_k}} F^{n_k} \eta_{\bar{v}}\}_{k \in \mathbb{N}}$ has a convergent subsequence which, without loss of generality, we assume to be the same sequence, i.e. there exists $\eta_\infty \in S_{\bar{v},c_+}$ such that

$$\eta_\infty = \lim_{k \to \infty} T^{-j_{n_k}} F^{n_k} \eta_{\bar{v}} .$$

[3] Helly's Selection Theorem states that if $\{f_n\}$ is a sequence of monotonically increasing functions on $\mathbb{R}$ with $0 \leq f_n(x) \leq 1$ for all x and n, then there is a function f and a sequence $\{n_k\}$ such that

$$f(x) = \lim_{k \to \infty} f_{n_k}(x) \quad \text{for every } x \in \mathbb{R} .$$

See Chap. 10 in [9] or exercise 13, Chap. 7 in [11].

Now we consider the sequence $T^{-m\bar{v}}F^m\eta_\infty$. Since $\eta_\infty \in S_{\bar{v},c_+}$, this sequence has the following property

$$\eta_{\bar{v}} \leq T^{-(m+1)\bar{v}}F^{m+1}\eta_\infty \leq T^{-m\bar{v}}F^m\eta_\infty \ .$$

Therefore it converges to a limit function

$$\phi = \lim_{m\to\infty} T^{-m\bar{v}}F^m\eta_\infty \ .$$

By continuity of F, the function $\phi \in \mathcal{I}$ satisfies the front equation $T^{\bar{v}}\phi = F\phi$ and is such that $\lim_{x\to+\infty}\phi(x) = 1$. However, one cannot conclude that $\lim_{x\to-\infty}\phi(x) = 0$ but only that

$$\lim_{x\to-\infty}\phi\,(x) \in \{0, c\} \ .$$

In order to complete the proof, one first shows that if f and h are such that

$$f'\,(c) = +\infty \text{ and } \inf\{x \in \mathbb{R} \ : \ h(x) > 0\} = -\infty$$

then $\lim_{x\to-\infty}\phi\,(x) = 0$ and the existence of fronts is proved in this case. In the general case, the conclusion follows

(1) by showing that any f (resp. h) can be approximated to arbitrary accuracy by local maps (resp. distribution functions) satisfying the previous assumptions and

(2) by showing that the limit of a sequence of extended bistable maps having fronts with converging velocities possesses itself a front (with the limit velocity).

3 Extended Circle Maps

3.1 Frenkel-Kontorova Models and Extended Circle Maps

One-dimensional chains of particles coupled by springs and placed in a periodic potential are represented by doubly infinite real sequences $\{u_s\}_{s\in\mathbb{Z}}$ (u_s represents the location of the s-th particle). In the dissipative limit, the dynamics of such chains, when driven by a constant force, are described by the gradient of a Frenkel-Kontorova (FK) functional (see the chapters by Floría, Baesens and Gómez-Gardeñes and by Baesens). It means that the sequences evolve according to the differential equation

$$\partial_t u_s = V'(u_s) + D + (u_{s-1} - 2u_s + u_{s+1})$$

where the potential V is periodic $V(x+1) = V(x)$ and $D \in \mathbb{R}$ is the driving force.

This equation is a special case of the following one

$$\partial_t u_s = -\left(g_2'(u_{s-1}, u_s) + g_1'(u_s, u_{s+1})\right) + D$$

where $g : \mathbb{R}^2 \to \mathbb{R}$ is a C^2 function satisfying the periodic condition

$$g(x+1, y+1) = g(x, y)$$

and such that the partial derivative $g_{12}''(x, y) \leq 0$ for all $(x, y) \in \mathbb{R}^2$ (twist condition).

This section concerns the dynamics of discrete time analogues of such equation; namely discrete time dynamical system defined by

$$u_s^{t+1} = u_s^t - \varepsilon\left(g_2'(u_{s-1}^t, u_s^t) + g_1'(u_s^t, u_{s+1}^t)\right) + \varepsilon D \tag{9}$$

where $t \in \mathbb{Z}$ and $\varepsilon > 0$ is the discretisation step. Special emphasis will be put on travelling wave solutions whose shape is an increasing periodic (in a suitable sense) function. Precisely, our concern will be with orbits given by

$$u_s^t = \psi\left(\alpha s + \nu_\alpha t\right) \tag{10}$$

where ψ is an increasing function such that $\psi\left(x+1\right) = \psi\left(x\right) + 1$. The number $\alpha > 0$ is called the mean spacing (wave number) of the wave and the number ν_α is called the rotation number (frequency).

A special case of travelling waves is when $\nu_\alpha = 0$ for which the configurations are stationary. In particular, according to relation (9), the existence of such stationary configuration for $D = 0$ is nothing else than the famous Aubry-Mather Theorem [8].

Just as done for bistable extended maps, we extend the analysis to systems with continuous space variable. That is to say, rather than considering sequences $u_s : \mathbb{Z} \to \mathbb{R}$, we consider functions $u(x) : \mathbb{R} \to \mathbb{R}$. In this larger phase space, the dynamics writes $u^{t+1} = Fu^t$ where

$$\left(Fu\right)\left(x\right) = u\left(x\right) - \varepsilon\left(g_2'(u\left(x-1\right), u\left(x\right)) + g_1'(u\left(x\right), u\left(x+1\right))\right) + \varepsilon D\,. \tag{11}$$

In order to deal with separate sets of travelling waves for distinct mean spacing, for each $\alpha > 0$, we consider the set (see Fig. 4)

$$\mathcal{N}_\alpha = \left\{u : u : \mathbb{R} \to \mathbb{R},\ u \text{ increasing and } u(x + \tfrac{1}{\alpha}) = u\left(x\right) + 1\right\}\,.$$

Since for any function ψ satisfying $\psi(x+1) = \psi(x) + 1$ the function $\phi\left(x\right) := \psi\left(\alpha x\right)$ belongs to $\mathcal{N}_\alpha$, the travelling wave solutions (10) can be written as follows

$$u^t = T^{-\frac{\nu_\alpha}{\alpha}t}\phi$$

where $\phi \in \mathcal{N}_\alpha$ and T^v is again the translation operator defined by $T^v u(x) = u(x - v)$.

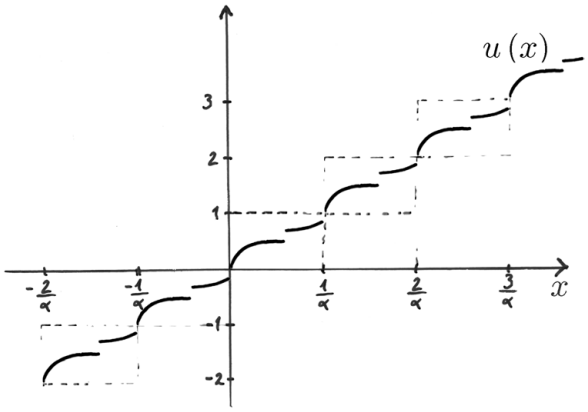


Fig. 4. A function u in $\mathcal{N}_\alpha$

Note that horizontal translations in $\mathcal{N}_\alpha$ can be viewed as vertical ones. So it is indifferent to view travelling waves either as propagating vertically or horizontally.

Just as bistable extended map do, the maps defined (11) commute with translations and are continuous with respect to pointwise convergence. In addition they can be shown to be increasing.

Lemma 3.1. [7] *For every $L > 0$, there exists $\varepsilon_L > 0$ such that, for every $\varepsilon \in (0, \varepsilon_L]$, $\alpha \in (0, L]$ and $u, v \in \mathcal{N}_\alpha$, we have $u \leq v$ implies $Fu \leq Fv$.*

At once, the map F satisfies the following properties

(a) For some $\alpha > 0$, or for all $\alpha > 0$, F maps $\mathcal{N}_\alpha$ into $\mathcal{N}_\alpha$.
(b) F is increasing, $u \leq v \Rightarrow Fu \leq Fv$.
(c) F is periodic, $F(u+1) = F(u) + 1$.
(d) F is homogeneous, $T^v F = FT^v \quad \forall v \in \mathbb{R}$.
(e) F is continuous, $\forall x \in \mathbb{R} \ \lim_{n\to\infty} u_n(x) = u(x) \Rightarrow \forall x \in \mathbb{R} \ \lim_{n\to\infty} Fu_n(x) = Fu(x)$.

The results in this section, in particular the uniqueness of rotation number (Proposition 3.1) and the existence of travelling waves (Theorem 3.1), hold for any map F which satisfies the properties a) to e). Such maps are called (spatially) **extended circle maps**.

The properties b), c) and d) imply that every set $\mathcal{N}_\alpha$ on which an extended circle map F is defined is invariant under the action of F. So the dynamics is well-defined.

The dynamics of functions with negative mean spacing (i.e. $\mathcal{N}_\alpha$ with $\alpha < 0$) is also included in this framework. Indeed, it suffices to apply the inversion $x \mapsto -x$ and to analyse the subsequent extended circle map. In the case $\alpha = 0$, the dynamics reduces to that of a lift of circle map (see below).

3.2 Coupled Lift of Circle Maps

A special example of extended circle map are the mappings which are formally identical to the bistable extended maps of the first section, namely

$$Fu = h * f \circ u \, . \tag{12}$$

However f is now a lift of a circle map (i.e. f is increasing, continuous map from $\mathbb{R}$ into itself for which $f(x+1) = f(x) + 1$ for all x) and h is now a distribution function satisfying $\int_{\mathbb{R}} |x| dh(x) < \infty$ [7]. Such maps cannot be interpreted as (extended) lift of coupled circle maps. Indeed for any integer function $n : \mathbb{R} \to \mathbb{N}$, we have $F(u+n) = F(u) + h * n$. But in general the function $h * n$ is not an integer function.

Anyway, coupled lift of circle maps (with discrete distribution function) can be interpreted as models of chains diffusively coupled particles in titled periodic potential (Frenkel-Kontorova models).

In the uncoupled case $h = H$ (where H is the Heaviside function) that is to say in the case where $Fu = f \circ u$, then Theorem 3.1 applied with $\alpha = 1$ states the existence of a semi-conjugacy to some translation for any lift of circle map f. Indeed, it states the existence of a lift of a circle map $\phi \in \mathcal{N}_1$ such that $f \circ \phi = T^{-\nu_1}\phi$ where ν_1 is the rotation number of F in $\mathcal{N}_1$ (see Proposition 3.1), the rotation number of f indeed.

3.3 Rotation Number of Extended Circle Maps

Unlike the analysis of fronts in bistable extended map for which the uniqueness of the velocity has been shown once a solution has been exhibited, the proof of existence of travelling waves with periodic shape begins with uniqueness of the rotation number.

In addition, the proof itself is simpler because the set $\mathcal{N}_\alpha$ containing the solutions are compact. This is also a reason why the proof extends to arbitrary circle maps F and holds not only for maps of the form $h * f \circ u$ (or their extension $\sum_k h_k * f_k \circ u$).

Horizontal displacements will be measured by using the reference zero:

$$J(u) := J_0(u) = \inf\{x \in \mathbb{R} \ : \ u(x) \geq 0\}$$

which is finite for every $u \in \mathcal{N}_\alpha$ with $\alpha > 0$ and satisfies the properties $J(T^\nu u) = J(u) + \nu$ and $u \leq v$ implies $J(u) \geq J(v)$.

The existence and the uniqueness of the rotation number for extended circle maps is given in the following statement.

Proposition 3.1. *Let F be an extended circle map defined on $\mathcal{N}_\alpha$ for some $\alpha > 0$. For every $u \in \mathcal{N}_\alpha$ and every $x \in \mathbb{R}$, the limit $\nu_\alpha := \lim_{t\to\infty} \frac{F^t u(x)}{t}$ exists and does not depend on x nor on u.*

Furthermore, we have $\left|J(F^t u) + \frac{\nu_\alpha t}{\alpha}\right| \leq \frac{2}{\alpha}$ *for all* $t \in \mathbb{N}$ *. Hence*

$$\nu_\alpha = \lim_{t\to\infty} \frac{F^t u(x)}{t} = -\,\alpha \lim_{t\to\infty} \frac{J(F^t u)}{t} \ .$$

The existence of the rotation number $\nu_\alpha = \lim_{t\to\infty} \frac{F^t u(x)}{t}$ extends to functions which need not be periodic but which can be sandwiched between two functions in $\mathcal{N}_\alpha$, see [7] for more details.

Proof: Every function $u \in \mathcal{N}_\alpha$ satisfies the inequalities

$$\varphi_\alpha^- \leq T^{-J(u)} u < \varphi_\alpha^+ \tag{13}$$

where $u < v$ means $u \leq v$ and $u \neq v$ and where the functions φ_α^- and φ_α^+ are defined by $\varphi_\alpha^-(x) = \lceil \alpha x \rceil - 1$ and $\varphi_\alpha^+ = \lfloor \alpha x \rfloor + 1$ for all $x \in \mathbb{R}$, see Fig. 5.

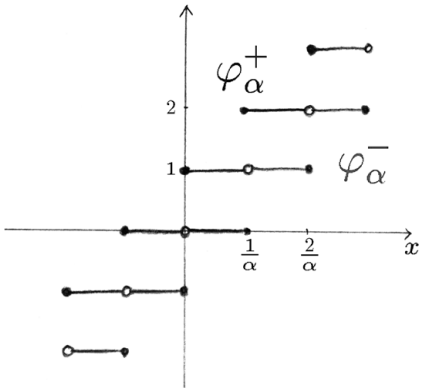


Fig. 5. The functions φ_α^- and φ_α^+

The quantity $j_t := J(F^t \varphi_\alpha^-)$ is finite for every $t \in \mathbb{N}$. The inequalities (13) imply $\varphi_\alpha^- \leq T^{-j_t} F^t \varphi_\alpha^-$ and $T^{-(j_t - \frac{1}{\alpha})} F^t \varphi_\alpha^+ < \varphi_\alpha^+$ because $J(F^t \varphi_\alpha^+) = j_t - \frac{1}{\alpha}$ for all t. Applying F^s, we obtain

$$F^s \varphi_\alpha^- \leq T^{-j_t} F^{t+s} \varphi_\alpha^- \quad \text{and} \quad T^{-(j_t - \frac{1}{\alpha})} F^{t+s} \varphi_\alpha^+ \leq F^s \varphi_\alpha^+$$

and then $j_s \geq j_{t+s} - j_t$ and $(j_{t+s} - \frac{1}{\alpha}) - (j_t - \frac{1}{\alpha}) \geq j_s - \frac{1}{\alpha}$. The sub-additivity of the sequence $\{j_t\}_{t\in\mathbb{N}}$ and the super-additivity of $\{j_t - \frac{1}{\alpha}\}_{t\in\mathbb{N}}$ imply that the following limit exists and is finite

$$\lim_{t\to\infty} \frac{j_t}{t} = \inf_{t>0} \frac{j_t}{t} = \sup_{t>0} \frac{j_t - \frac{1}{\alpha}}{t} \ .$$

We denote this quantity by $-\frac{\nu_\alpha}{\alpha}$. A consequence is the following important inequality

$$-\frac{\nu_\alpha}{\alpha}t \le j_t \le -\frac{\nu_\alpha}{\alpha}t + \frac{1}{\alpha}\,, \quad \forall t \in \mathbb{N}\,.$$

We are now about to prove the existence and uniqueness of the rotation number for the function φ_α^-. By applying the inequalities (13) to the function $F^t\varphi_\alpha^-$, we obtain

$$\varphi_\alpha^- - \lceil j_t\alpha \rceil = T^{\frac{1}{\alpha}\lceil j_t\alpha\rceil}\varphi_\alpha^- \le F^t\varphi_\alpha^- < T^{\frac{1}{\alpha}\lfloor j_t\alpha\rfloor}\varphi_\alpha^+ = \varphi_\alpha^+ - \lfloor j_t\alpha \rfloor\,.$$

As a consequence, for every $x \in \mathbb{R}$ we have

$$\nu_\alpha = -\lim_{t\to\infty}\frac{\lfloor j_t\alpha\rfloor}{t} \le \liminf_{t\to\infty}\frac{F^t\varphi_\alpha^-(x)}{t} \le \limsup_{t\to\infty}\frac{F^t\varphi_\alpha^-(x)}{t} \le -\lim_{t\to\infty}\frac{\lceil j_t\alpha\rceil}{t} = \nu_\alpha$$

the rotation number associated with φ_α^- exists and does not depend on x. In addition, $\varphi_\alpha^- + 1 \le \varphi_\alpha^+ \le \varphi_\alpha^- + 2$ and thus the rotation number also exists and does not depend on x for the function φ_α^+. Finally, by applying the inequalities (13) we conclude the same results for any $u \in \mathcal{N}_\alpha$ □

3.4 Existence of Travelling Waves

Proposition 3.1 claimed that every configuration in $\mathcal{N}_\alpha$ propagates asymptotically with a unique (horizontal) velocity. The main theorem below claims the existence of a configuration which the action of F amounts to a translation by $-\nu_\alpha/\alpha$.

Theorem 3.1. [7] *Let F be an extended circle map defined on $\mathcal{N}_\alpha$ for some $\alpha > 0$. There exists $\phi \in \mathcal{N}_\alpha$ such that $F\phi = T^{-\frac{\nu_\alpha}{\alpha}}\phi$.*

The proof is similar to the proof of front existence. As said before, the main difference resides in the compactness of $\mathcal{N}_\alpha$ which considerably simplifies the proof.

In a first step, we consider the set of sub-solutions of the travelling wave equation. Given $\nu \in \mathbb{R}$, we define

$$\mathcal{S}(\nu) = \left\{u \in \mathcal{N}_\alpha \;:\; T^{\frac{\nu}{\alpha}}Fu \le u \text{ and } J(u) = 0\right\}\,.$$

Next we show the rotation number can be defined using these sets:

Lemma 3.2. $\nu_\alpha = \inf\{\nu \in \mathbb{R} \;:\; \mathcal{S}(\nu) \neq \emptyset\}$.

Proof of the Lemma: Given $t \in \mathbb{N}$, let the function φ_t be defined by

$$\varphi_t(x) = \min_{0\le s<t}\left\{T^{-\frac{s}{t}(j_t-\frac{1}{\alpha})}F^s\varphi_\alpha^+(x)\right\}, \quad \forall x \in \mathbb{R}$$

where j_t and φ_α^+ were introduced in the proof of Proposition 3.1. The fact that $\varphi_\alpha^+ \in \mathcal{N}_\alpha$ and the properties of F ensure that $\varphi_t \in \mathcal{N}_\alpha$ for every t. Thus all $J(\varphi_t)$ are finite. Moreover, by monotony of F, we have $T^{-\frac{1}{t}(j_t-\frac{1}{\alpha})}F\varphi_t \le T^{-\frac{s+1}{t}(j_t-\frac{1}{\alpha})}F^{s+1}\varphi_\alpha^+$ for every $0 \le s < t$. This implies that

$$T^{-\frac{1}{t}(j_t-\frac{1}{\alpha})}F\varphi_t \leq \min_{1\leq s\leq t} T^{-\frac{s}{t}(j_t-\frac{1}{\alpha})}F^s\varphi_\alpha^+ \leq \min_{0\leq s<t} T^{-\frac{s}{t}(j_t-\frac{1}{\alpha})}F^s\varphi_\alpha^+ = \varphi_t$$

because $T^{-(j_t-\frac{1}{\alpha})}F^t\varphi_\alpha^+ \leq \varphi_\alpha^+$ as indicated by the right inequality (13). We have shown that the set $\mathcal{S}(-\frac{\alpha}{t}(j_t-\frac{1}{\alpha}))$ is not empty for every $t>0$. Therefore, we have

$$\nu_\alpha = -\alpha \lim_{t\to\infty} \frac{1}{t}\left(j_t - \frac{1}{\alpha}\right) \geq \inf\left\{\nu\in\mathbb{R} \ : \ \mathcal{S}(\nu)\neq\emptyset\right\}.$$

On the other hand, we assume that $u\in\mathcal{S}(\nu)\neq\emptyset$ for some $\nu\in\mathbb{R}$. Then $\varphi_\alpha^- \leq u$ by relation (13) and thus $F^t\varphi_\alpha^- \leq F^t u \leq T^{-\frac{\nu}{\alpha}t}u$ which implies $j_t \geq -\frac{\nu}{\alpha}t$, i.e. $\nu \geq -\alpha\frac{j_t}{t}$ for all $t>0$. Consequently, we have

$$\inf\left\{\nu\in\mathbb{R} \ : \ \mathcal{S}(\nu)\neq\emptyset\right\} \geq \nu_\alpha \ .$$

□

As in the proof of front existence, the second step consists of considering a minimal sub-solutions, namely we consider the function

$$\eta_\nu(x) = \inf_{u\in\mathcal{S}(\nu)} u(x), \quad \forall x\in\mathbb{R} \ .$$

It turns out that $\mathcal{S}(\nu_\alpha)\neq\emptyset$ and $\eta_{\nu_\alpha}\in\mathcal{S}(\nu_\alpha)$.

In a third step, we consider the sequence $\left\{T^{n\frac{\nu_\alpha}{\alpha}}F^n\eta_{\nu_\alpha}\right\}_{n\in\mathbb{N}}$. By monotony and homogeneity, we have

$$T^{(n+1)\frac{\nu_\alpha}{\alpha}}F^{n+1}\eta_{\nu_\alpha} \leq T^{n\frac{\nu_\alpha}{\alpha}}F^n\eta_{\nu_\alpha}, \quad \forall n\in\mathbb{N} \ .$$

In addition, one can show that $T^{\frac{1}{\alpha}}\varphi_\alpha^- \leq T^{n\frac{\nu_\alpha}{\alpha}}F^n\eta_{\nu_\alpha}$ and hence that the sequence is bounded from below. Consequently, this sequence converges pointwise to the limit function $\phi\in\mathcal{N}_\alpha$ which satisfies $F\phi = T^{-\nu_\alpha/\alpha}\phi$. We refer to [7] for more details.

3.5 Continuity of the Rotation Number

Just as the front velocity, the rotation number ν_α varies continuously with changes of extended circle maps (and in particular with their parameters).

For coupled lift of circle maps $Fu = h * f\circ u$, changes are the same as before; namely pointwise convergence for the local map f and convergence in the Hausdorff topology for the distribution function h (an additional condition is needed to ensure that the distribution functions h satisfy $\int_{\mathbb{R}}|x|\,dh(x)<\infty$, see Lemma 3.3 in [7]).

For the discrete time version (9) of the Frenkel-Kontorova model, a continuous dependence with parameters and with the generating function has been shown [7]. All these results are deduced from the following statement valid for arbitrary extended circle maps. If $\lim_{n\to\infty}\sup_{u\in\mathcal{N}_\alpha} d(F_n(u),F(u))=0$ where the distance $d(\cdot,\cdot)$ is the Hausdorff distance, then $\lim_{n\to\infty}\nu_\alpha(F_n)=\nu_\alpha(F)$.

In complement to continuity with respect to changes in the map, the rotation number depends continuously on the mean spacing α. This is proved by using bigger spaces $\mathcal{M}_{\alpha',\alpha''}$ which contain $\mathcal{N}_\alpha$ for every $\alpha'\leq\alpha\leq\alpha''$.

3.6 Extended Circle Maps with Vanishing Rotation Number

Recall that Aubry-Mather Theorem states that any Frenkel-Kontorova functional possess stationary configurations for every mean spacing α [8].

In the present framework, this amounts to say that the map F defined by (11) with $D = 0$ has, for every α, a fixed point in $\mathcal{N}_\alpha$. By Proposition 3.1 and Theorem 3.1, this result would follow from the fact that $\nu_\alpha = 0$ for all α.

It turns out that the property $\nu_\alpha = 0$ for all $\alpha > 0$ is not limited to the model (11) with $D = 0$. As stated in the next statement, it extends to any extended circle map satisfying some symmetry condition.

Theorem 3.2. *If there exists a lift of a circle map $\tilde{f}$ such that the following relation holds*

$$\int_0^{\frac{1}{\alpha}} (Fu - u) d(\tilde{f} \circ u) = 0 \tag{14}$$

for every continuous function $u \in \mathcal{N}_\alpha$ ($\alpha > 0$), then the rotation number $\nu_\alpha = 0$.

The proof essentially relies on various properties of the Lebesgue-Stieltjes integral [7]. The fact that the model (11) with $D = 0$ satisfies this property, however, is elementary and claim in our final statement.

Proposition 3.2. *For every generating function g, every $\alpha > 0$, and every $\varepsilon > 0$, the map*

$$F_\varepsilon u(x) = u(x) - \varepsilon \left(g_2'(u(x-1), u(x)) + g_1'(u(x), u(x+1)) \right) \quad \forall x \in \mathbb{R} \,,$$

satisfies the condition (14) with $\tilde{f}(x) = x$.

Indeed by using $T^{-\frac{1}{\alpha}} u = u + 1$, we have

$$\begin{aligned} &\int_0^{\frac{1}{\alpha}} (F_\varepsilon u - u) du \\ &\quad = -\varepsilon \int_0^{\frac{1}{\alpha}} g_2'(u(x), u(x+1)) du(x+1) - \varepsilon \int_0^{\frac{1}{\alpha}} g_1'(u(x), u(x+1)) du(x) \\ &\quad = -\varepsilon \int_0^{\frac{1}{\alpha}} dg(u(x), u(x+1)) = 0 \end{aligned}$$

for every continuous function $u \in \mathcal{N}_\alpha$.

References

1. R. Carretero-González, D.K. Arrowsmith and F. Vivaldi, *One-dimensional dynamics for traveling fronts in coupled map lattices*, Phys. Rev. E **61** (2000) 1329–1336. 267

2. C. Baesens and R.S. MacKay, *Gradient dynamics of tilted Frenkel-Kontorova models*, Nonlinearity **11** (1998) 949–964.
3. R. Coutinho and B. Fernandez, *Extended symbolic dynamics in bistable CML: Existence and stability of fronts*, Physica D **108** (1997) 60–80. 265, 267
4. R. Coutinho and B. Fernandez, *On the global orbits in a bistable CML*, Chaos **7** (1997) 301–310. 265
5. R. Coutinho and B. Fernandez, *Fronts and interfaces in bistable extended mappings*, Nonlinearity **11** (1998) 1407–1433. 265
6. R. Coutinho and B. Fernandez, *Fronts in extended systems of bistable maps coupled by convolutions*, Nonlinearity **17** (2004) 23–47. 265, 272, 275
7. R. Coutinho and B. Fernandez, *Spatially extended circle maps: Monotone periodic dynamics of functions with linear growth*, J. Nonlinear Sci. **14** (2004) 93–118. 265, 278, 279, 280, 281, 282, 283
8. C. Golé, *Symplectic twist maps*, World Scientific (2001). 277, 283
9. A. Kolmogorov and S. Fomin, *Elements of the theory of functions and of functional analysis*, Mir (1976). 275
10. E. Lukacs, *Characteristic functions*, Hafner Publishing Co. (1970). 272
11. W. Rudin, *Principles of mathematical analysis, 3rd edition*, McGraw-Hill (1976). 275
12. H. Weinberger, *Long-time behaviour of a class of biological models*, SIAM J. Math. Anal. **13** (1982) 353–396. 265

Spatially Extended Systems with Monotone Dynamics (Continuous Time)

C. Baesens

Mathematics Institute, University of Warwick. Coventry CV4 7AL, UK
claude@maths.warwick.ac.uk

1 Introduction

Broadly speaking, *monotone* dynamics means that some partial order is preserved under the dynamics, that is, if two solutions of some differential equation are ordered at an initial time, they remain in the same order at later times; we also speak of *order-preserving* dynamics. This monotonicity property makes the dynamics comparatively simple.

In the theory of partial differential equations, monotonicity methods are well-established, based on maximum principles. Maximum principles are fundamental tools to compare solutions of initial boundary value problems associated to systems of (non-linear) parabolic PDEs, and to establish existence of positively invariant sets and monotonicity properties of the dynamics. An example of this type of result is that generically a bounded solution for a reaction-diffusion system converges asymptotically to an equilibrium. For an overview of maximum principles for PDEs we refer to the textbook by Protter and Weinberger [26], and for the main monotonicity results on reaction-diffusion equations we refer to the monograph by Smith on Monotone Dynamical Systems [28] and the bibliography therein (up to 1995).

In ordinary differential equations, monotonicity methods and comparison arguments go back to the 1930's with the works of Müller [25] and Kamke [18] but really took off in the dynamical systems community in the 1980's in particular with the important contributions of Hirsch, e.g. [14, 15] for so-called cooperative systems of ODEs, leading to what is now described as monotone dynamical systems theory. Strict cooperativity for coupled ODEs is a condition on the interactions between variables which plays the same role as the diffusion term in reaction-diffusion equations and is responsible for the monotonicity of the dynamics. For an overview and references on monotone dynamical systems we refer again to [28]. Most results reported there concern bounded solutions. For example, as for parabolic PDEs, under some conditions on the partial order, generic bounded solutions of strictly cooperative systems of ODEs converge asymptotically to equilibria.

C. Baesens: *Spatially Extended Systems with Monotone Dynamics (Continuous Time)*, Lect. Notes Phys. **671**, 241–263 (2005)
www.springerlink.com

In this chapter we shall concentrate on monotone dynamics in networks of coupled ODEs. In the first part, we shall consider cooperative systems of first order ODEs. We shall first show how cooperativity of the interaction between sites implies monotonicity of the dynamics and how strict cooperativity leads to strict monotonicity. An example of strictly cooperative extended systems are Frenkel-Kontorova (FK) models with gradient dynamics. For such models we shall use strict monotonicity to prove the existence of rotationally ordered spatially periodic equilibria for any mean spacing, providing an alternative proof for a result from Aubry-Mather theory of minimum energy states. We shall also establish that spatially periodic solutions converge asymptotically to equilibria if there are any. If there are no such equilibria, as is the case for FK models with strong enough additive forcing term, then solutions are unbounded and strict monotonicity will be used to prove that they converge asymptotically to a unique periodically sliding solution.

Mechanical systems satisfy Newton's second law so are described by second order differential equations. For such systems gradient dynamics can been seen as a strong damping limit so that the inertial term is negligible, thereby reducing second order ODEs to first order ODEs. In the second part of this chapter, we shall consider networks of damped mechanical units with cooperative coupling and we shall show that if the damping is strong enough (overdamping) then there is again a partial order which is preserved under the dynamics and strict monotonicity if the system is strictly cooperative. Our example of application will be the overdamped dynamics of FK models for which strict monotonicity will allow us to prove similar results as for the gradient dynamics.

2 First Order Local Dynamics

In this section we consider networks of coupled units with one-dimensional local dynamics, so the setting is a system of first order coupled ODEs:

$$\dot{x}_n = f_n(\boldsymbol{x}, t) \tag{1}$$

for n in some index set $\mathcal{S}$, and $\boldsymbol{x} := (x_j)_{j\in\mathcal{S}} \in \mathbb{R}^{\mathcal{S}}$; x_j represents the *state of unit* j, $\boldsymbol{x}$ denotes the *state of the network* and the *state space* X will be $\mathbb{R}^{\mathcal{S}}$ or some subset of $\mathbb{R}^{\mathcal{S}}$. The set of sites $\mathcal{S}$ may be finite or infinite, for instance $\mathcal{S} = \mathbb{Z}^d$ for some d. We shall assume that the function $\boldsymbol{f} := (f_j)_{j\in\mathcal{S}} : X \times \mathbb{R}$ is continuously differentiable with respect to uniform topology. In such case the solution $\boldsymbol{x}(t)$ (alternatively denoted $\phi_t(\boldsymbol{x}(0))$) to the initial value problem exists and is unique for any initial value $\boldsymbol{x}(0)$ and times $|t|$ small enough. For many physically motivated systems the solutions exist for all positive times. If the system (1) is autonomous and the solutions are defined for all times $t \in \mathcal{T} = \mathbb{R}$ or $\mathbb{R}_+$, then the map $\phi : \mathcal{T} \times X \to X : (t, \boldsymbol{x}) \to \phi_t(\boldsymbol{x})$ is called a (global) *flow* or *semi-flow* respectively.

Cooperativity Condition

We shall assume that the functions $f_n : \mathbb{R}^{\mathcal{S}} \times \mathbb{R}_+ \to \mathbb{R}$ satisfy

$$\frac{\partial f_n}{\partial x_j}(\boldsymbol{x}, t) \geq 0 \quad \forall j \neq n \qquad \text{(cooperativity condition)} \tag{2}$$

in the domain of interest, that is, the derivative of the forcing by a neighbour with respect to the state of the neighbour is non-negative. Systems of the form (1) satisfying condition (2) are called *cooperative systems*, a terminology which was introduced by the biology community studying population dynamics and then brought to the dynamical systems community by Hirsch.

Note that system (1) is called a *competitive system* if the inequalities in (2) are reversed: $\frac{\partial f_n}{\partial x_j}(\boldsymbol{x}, t) \leq 0$ for all $j \neq n$, thus, by time reversal, a competitive system becomes cooperative and vice-versa.

2.1 Cooperativity Implies Monotonicity

We use here the natural partial order *"less than or equal to"* on sequences in $\mathbb{R}^{\mathcal{S}}$, defined by

$$\boldsymbol{x} \leq \boldsymbol{y} \quad \text{if, for all } i \in \mathcal{S}, \;\; x_i \leq y_i \; . \tag{3}$$

We use also the following additional notation.

$$\boldsymbol{x} \geq \boldsymbol{y} \quad \text{if} \quad \boldsymbol{y} \leq \boldsymbol{x} \qquad \text{"greater than or equal to"} \tag{4}$$

$$\boldsymbol{x} < \boldsymbol{y} \quad \text{if} \quad \boldsymbol{x} \leq \boldsymbol{y} \text{ and } \boldsymbol{x} \neq \boldsymbol{y} \qquad \text{"less than"} \tag{5}$$

$$\boldsymbol{x} > \boldsymbol{y} \quad \text{if} \quad \boldsymbol{y} < \boldsymbol{x} \qquad \text{"greater than"} \; . \tag{6}$$

We say that the dynamics are *monotone* or *order-preserving* if for any $\boldsymbol{x}(0), \boldsymbol{y}(0) \in \mathbb{R}^{\mathcal{S}}$

$$\boldsymbol{x}(0) \leq \boldsymbol{y}(0) \implies \boldsymbol{x}(t) \leq \boldsymbol{y}(t) \tag{7}$$

for all $t > 0$ for which both are defined. So if two initial points are ordered, the solutions starting from these points remain in the same order for all positive times.

We now show that the dynamics of cooperative systems (1) are monotone, i.e. preserve the partial order (3).

Proposition 2.1 (Weak monotonicity of the dynamics). *If system (1) is cooperative, i.e. satisfies condition (2), then* $\boldsymbol{x}(0) \leq \boldsymbol{y}(0)$ *implies that* $\boldsymbol{x}(t) \leq \boldsymbol{y}(t)$ *for all positive* t *such that both are defined.*

Proof. We first prove the monotonicity of the *linearised* dynamics about any path $\boldsymbol{x}(t)$. Then we show how the monotonicity follows for the full system (1) by taking a homotopy between the two solutions $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$.[1]

[1] This strategy is similar to that in parabolic PDEs: maximum principles are established for linear equations but can be used to compare solutions of the nonlinear equations because the mean value theorem produces a linear equation for the difference of two solutions.

(i) The linearisation of system (1) about a path $\boldsymbol{x}(t)$ takes the form

$$\dot{\xi}_n = \sum_{j\in\mathcal{S}} \frac{\partial f_n}{\partial x_j}(\boldsymbol{x}(t))\,\xi_j\,, \qquad n \in \mathcal{S} \tag{8}$$

or, in compact form,

$$\dot{\boldsymbol{\xi}} = A(\boldsymbol{x}(t))\,\boldsymbol{\xi}\,.$$

The non-diagonal entries of the matrix $A(\boldsymbol{x}(t))$ are non-negative by the cooperativity condition. To take care of the diagonal terms, we set $\mu = \sup_{k,\boldsymbol{x}} -\frac{\partial f_k}{\partial x_k}(\boldsymbol{x})$ (assuming for instance that the $\frac{\partial f_k}{\partial x_k}$ are uniformly bounded in the range of interest), and define $\boldsymbol{\gamma}(t) := e^{\mu t}\boldsymbol{\xi}(t)$. Then $\boldsymbol{\gamma}(t) \geq \mathbf{0}$ if and only if $\boldsymbol{\xi}(t) \geq \mathbf{0}$, and $\boldsymbol{\gamma}$ satisfies the linear system

$$\dot{\gamma}_k = \left(\mu + \frac{\partial f_k}{\partial x_k}\right)\gamma_k + \sum_{j\neq k} \frac{\partial f_k}{\partial x_j}\gamma_j\,, \qquad k \in \mathcal{S} \tag{9}$$

or

$$\dot{\boldsymbol{\gamma}} = B(\boldsymbol{x}(t))\boldsymbol{\gamma}$$

where the matrix $B(\boldsymbol{x}(t)) = A(\boldsymbol{x}(t)) + \mu Id$ is non-negative definite (for any $\boldsymbol{x}$ and t). It then follows that $\boldsymbol{\gamma}(t) \geq \mathbf{0}$ for $t > 0$ if $\boldsymbol{\gamma}(0) \geq \mathbf{0}$, which can be seen for instance using Picard iterations:

$$\boldsymbol{\gamma}^{(n+1)}(t) = \boldsymbol{\gamma}(0) + \int_0^t B(\boldsymbol{x}(s))\boldsymbol{\gamma}^{(n)}(s)\,ds, \quad \boldsymbol{\gamma}^{(0)}(t) = \boldsymbol{\gamma}(0)\,.$$

Hence $\boldsymbol{\xi}(t) = e^{-\mu t}\boldsymbol{\gamma}(t) \geq \mathbf{0}$ for positive t if $\boldsymbol{\xi}(0) = \gamma(0) \geq \mathbf{0}$.

(ii) Let $\boldsymbol{x}^\lambda(0) := (1-\lambda)\boldsymbol{x}(0) + \lambda\boldsymbol{y}(0)$ for $\lambda \in [0,1]$, and let $\boldsymbol{x}^\lambda(t)$ be the solution of (1) with initial condition $\boldsymbol{x}^\lambda(0)$. Define $\boldsymbol{\xi}^\lambda(t) := \frac{\partial}{\partial\lambda}\boldsymbol{x}^\lambda(t)$. Then $\boldsymbol{\xi}^\lambda$ satisfies the linearised equation (8) about $\boldsymbol{x}^\lambda$(t), and $\boldsymbol{\xi}^\lambda(0) = \boldsymbol{y}(0) - \boldsymbol{x}(0) \geq \mathbf{0}$. It then follows from part (i) that $\boldsymbol{\xi}^\lambda(t) \geq \mathbf{0}$ for any positive t and hence

$$\boldsymbol{y}(t) = \boldsymbol{x}(t) + \int_0^1 \boldsymbol{\xi}^\lambda(t)\,d\lambda \geq \boldsymbol{x}(t) \tag{10}$$

as required, where $\int_0^1 \boldsymbol{\xi}^\lambda(t)\,d\lambda$ denotes the vector in $\mathbb{R}^{\mathcal{S}}$ with n-th component $\int_0^1 \xi_n^\lambda(t)\,d\lambda$. □

2.2 Strict Cooperativity Implies Strict Monotonicity

We define a relation $\ll$ "strictly less than" on $\mathbb{R}^{\mathcal{S}}$ by

$$\boldsymbol{x} \ll \boldsymbol{y} \quad \text{if } x_n < y_n \text{ for all } n \in \mathcal{S}\,. \tag{11}$$

We say that the dynamics are *strictly monotone* if for any $\boldsymbol{x}(0)$, $\boldsymbol{y}(0) \in \mathbb{R}^{\mathcal{S}}$

$$\boldsymbol{x}(0) < \boldsymbol{y}(0) \implies \boldsymbol{x}(t) \ll \boldsymbol{y}(t) \tag{12}$$

for all $t > 0$ for which both are defined. So if two distinct initial points are ordered, all components of the solutions starting from these points are strictly ordered for all positive times. This means that it is enough for one component $y_n - x_n$ to be positive at $t = 0$ to make all the components of $\boldsymbol{y}$ and $\boldsymbol{x}$ be pulled apart at later times. In the case of a finite network, our definition is equivalent to those of *strongly order-preserving* [23] and *strongly monotone* [16]. For infinite networks, all three are different [28].

With the following slightly stronger assumption, we shall show that the dynamics of system (1) are strictly monotone.

We say that system (1) is *strictly cooperative* if for each pair $i, n \in \mathcal{S}$ there is a j-chain $n = n_0, n_1, \ldots, n_j = i$ for some $j \geq 1$, $n_k \neq n_{k+1}$, for all k, and for $k = 0, \ldots, j-1$ there is c_k such that

$$\frac{\partial f_{n_k}}{\partial x_{n_{k+1}}}(\boldsymbol{x}) \geq c_k > 0 \qquad \text{for all } \boldsymbol{x} \in X\ . \tag{13}$$

Proposition 2.2 (Strict monotonicity). *If system (1) is strictly cooperative, then* $\boldsymbol{x}(0) < \boldsymbol{y}(0)$ *implies that* $\boldsymbol{x}(t) \ll \boldsymbol{y}(t)$ *for all* $t > 0$ *for which both are defined.*

Proof. Again, strict monotonicity is first obtained for the linearized dynamics, i.e. we show that $\boldsymbol{\xi}(0) > \mathbf{0}$ implies that $\boldsymbol{\xi}(t) \gg \mathbf{0}$ for all positive t, and then homotopy provides the result.

(i) By assumption $\boldsymbol{\xi}(0) > \mathbf{0}$, so there is a site i for which $\xi_i(0) > 0$, say $\xi_i(0) = \delta > 0$, and some k for which $\frac{\partial f_k}{\partial x_i} \geq c_{ki} > 0$. Then (9) gives $\dot{\gamma}_k \geq c_{ki}\gamma_i \geq c_{ki}\delta$ since $\dot{\gamma}_i \geq 0$ and $\gamma_i(0) = \xi_i(0) = \delta$. Therefore $\gamma_k(t) \geq c_{ki}\delta t$, hence

$$\xi_k(t) \geq c_{ki}\delta t e^{-\mu t}\ . \tag{14}$$

Continuing in such a way, for a site $n \in \mathcal{S}$ connected to site i by a j-chain $n = n_0, n_1, \ldots, n_j = i$ with $\frac{\partial f_{n_{k-1}}}{\partial x_{n_k}} \geq c_{n_{k-1}n_k} > 0$, $k = 1, \ldots, j$, we obtain

$$\xi_n(t) \geq \prod_{k=0}^{j-1} c_{n_k n_{k+1}} \delta e^{-\mu t} \frac{t^j}{j!} > 0\ .$$

(ii) Introducing $\boldsymbol{x}^\lambda$, $\lambda \in [0,1]$, and $\boldsymbol{\xi}^\lambda$ as in the proof of weak monotonicity, it follows from part (i) that $\boldsymbol{\xi}^\lambda(t) \gg \mathbf{0}$ for all $t > 0$ since $\boldsymbol{\xi}^\lambda(0) = \boldsymbol{y} - \boldsymbol{x} > \mathbf{0}$ by assumption. Hence

$$\boldsymbol{y}(t) = \boldsymbol{x}(t) + \int_0^1 \boldsymbol{\xi}^\lambda(t) d\lambda \gg \boldsymbol{x}(t)\ . \qquad \square \tag{15}$$

Now that we have established the monotonity/strict monotonicity of the dynamics of cooperative/strictly cooperative systems, we shall present some consequences for a particular class of extended systems, the Frenkel-Kontorova models.

3 Gradient Dynamics of the Frenkel-Kontorova Model

The Frenkel-Kontorova model can be seen as a doubly infinite one-dimensional chain of identical classical particles with convex nearest neighbour interaction, subject to a spatially periodic potential. The potential energy of the chain is the formal sum

$$H(\boldsymbol{x}) = \sum_{n\in\mathbb{Z}} h(x_n, x_{n+1}) = \sum_{n\in\mathbb{Z}} W(x_{n+1} - x_n) + V(x_n) \tag{16}$$

where $x_n \in \mathbb{R}$ denotes the position of particle n, $\boldsymbol{x}$ denotes $(x_n)_{n\in\mathbb{Z}}$, the function $h : \mathbb{R}^2 \to \mathbb{R}$ is C^2 with $h(x, x') = W(x' - x) + V(x)$; V is the periodic on-site potential, which we shall assume has period 1: $V(x + 1) = V(x)$, so that $h(x + 1, x' + 1) = h(x, x')$, and W is the interaction potential between (nearest) neighbours, which we shall assume is (uniformly) strictly convex: $W''(x) \geq b$ for some positive constant b. The potential energy of the chain is typically infinite but its gradient is well-defined. The Newtonian dynamics of such a chain is described by a system of second order ODEs

$$m\ddot{x}_n + \gamma\dot{x}_n = -\frac{\partial H}{\partial x_n}(\boldsymbol{x}), \qquad n \in \mathbb{Z} \tag{17}$$

where m denotes the mass of the particles, and γ is the damping factor. For many physical systems modelled by FK chains, for instance charge density wave materials and Josephson Junction arrays, the relevant parameter ranges are strong damping compared to inertia (see the chapter by Floría et al. in this book), so a first approximation is to drop the inertial term $m\ddot{x}_n$ in (17). In this section, we shall consider this limiting case with no inertial term, so that the local dynamics at each site are described by first order rather than second order ODEs; in a subsequent section we will see that keeping a small inertial term (overdamped second order local dynamics) leads to same qualitative results because the dynamics are order-preserving in the range too.

So (taking $\gamma = 1$ without loss of generality) we consider here what we call the *gradient dynamics*[2] of the Frenkel-Kontorova model,

$$\dot{x}_n = -\frac{\partial H}{\partial x_n}(\boldsymbol{x}), \qquad n \in \mathbb{Z} \tag{18}$$

which is a first order system of the form (1) with $f_n(\boldsymbol{x}, t) = -\frac{\partial H}{\partial x_n}(\boldsymbol{x})$ and $\mathcal{S} = \mathbb{Z}$. Equation (18) take the form

$$\dot{x}_n = W'(x_{n+1} - x_n) - W'(x_n - x_{n-1}) - V'(x_n), \qquad n \in \mathbb{Z}\,. \tag{19}$$

For instance, for the Standard FK model where the interaction between neighbours is simply harmonic ($W(x) = \frac{1}{2}x^2$) and the on-site potential is $V(x) = \frac{k}{4\pi^2}(1 - \cos 2\pi x)$, the latter equations become

[2] Some other authors use the terms "dissipative dynamics" [11] or "overdamped limit" [8] for this limit — see also the chapter by Floría et al. in this book.

$$\dot{x}_n = x_{n+1} - 2x_n + x_{n-1} - \frac{k}{2\pi} \sin 2\pi x_n \ . \tag{20}$$

The first three terms of the right hand side of (20) are just the discrete Laplacian, so that (20) is a spatially discrete version of a reaction-diffusion equation, as noticed by Angenent [1] who was the first to use monotonicity techniques for FK models.

System (19) is strictly cooperative because of the strict convexity of the interaction between neighbours, since the only non-zero non-diagonal terms of the derivative of f are

$$\frac{\partial f_n}{\partial x_{n-1}} = W''(x_n - x_{n-1}) \geq b \ , \qquad \frac{\partial f_n}{\partial x_{n+1}} = W''(x_{n+1} - x_n) \geq b \tag{21}$$

and given site $i > n$, take the chain $n, n+1, \ldots, i$ (similarly for $i < n$) and for $i = n$ take $i, i+1, i$.

So it follows from the previous section that gradient dynamics of FK models are strictly monotone. This also remains true for additively or parametrically driven FK models where

$$h(x, x') = C(t)W(x' - x) + K(t)V(x) - F(t)x \tag{22}$$

as long as $C(t) \geq 0$ for all t.[3] See the chapter by Floría et al. in this book for examples of time-dependent forcing where the monotonicity property is exploited.

We shall now derive a number of consequences of the monotonicity property of the dynamics.

The local flow ϕ_t of (19) commutes with the group of translations $\{\sigma_{ij} : (i, j) \in \mathbb{Z}^2\}$ defined by

$$(\sigma_{ij}\boldsymbol{x})_n = x_{n+i} + j \quad \text{for all } n \in \mathbb{Z} \ . \tag{23}$$

Thus if $\boldsymbol{x}$ is a solution of (19), so is any of its translates $\sigma_{ij}\boldsymbol{x}$. It follows that if $\sigma_{ij}\boldsymbol{x}(0) \leq \boldsymbol{x}(0)$ for some i, j, then the same holds at all future times for which $\boldsymbol{x}(t)$ is defined.

For $w \in \mathbb{N}$, say $\boldsymbol{x}$ has *width* $\leq w$ if $\forall m \in \mathbb{Z}$ there exists $n_-(m)$, $n_+(m) \in \mathbb{Z}$ with $0 \leq n_+(m) - n_-(m) \leq w$ such that

$$\sigma_{m,n_-(m)}\boldsymbol{x} \leq \boldsymbol{x} \leq \sigma_{m,n_+(m)}\boldsymbol{x} \ . \tag{24}$$

Then width $\leq w$ is preserved by the forward dynamics. If $\boldsymbol{x}$ has bounded width, then it has a *mean spacing* ρ, i.e.

$$\frac{x_n - x_m}{n - m} \to \rho \, , \qquad \text{as } n - m \to \infty \ . \tag{25}$$

In fact $|(x_n - x_m) - (n - m)\rho| \leq w$ [5]. The mean spacing is preserved under the forward dynamics (because $n_\pm(m)$ are).

[3] If $C(t) = 0$ for some t then the system is cooperative but not strictly cooperative.

Say $\boldsymbol{x} \in \mathbb{R}^{\mathbb{Z}}$ is *weakly rotationally ordered* (abbreviated WRO) if, for all $i, j \in \mathbb{Z}$,

$$\text{either} \qquad \sigma_{ij}\boldsymbol{x} \leq \boldsymbol{x} \qquad \text{or} \qquad \sigma_{ij}\boldsymbol{x} \geq \boldsymbol{x} \; . \tag{26}$$

(Equivalently $\boldsymbol{x}$ has width ≤ 1.) Similarly, say $\boldsymbol{x} \in \mathbb{R}^{\mathbb{Z}}$ is *(strictly) rotationally ordered* (abbreviated RO)[4] if, for all $i, j \in \mathbb{Z}$,

$$\text{either} \quad \sigma_{ij}\boldsymbol{x} \ll \boldsymbol{x} \quad \text{or} \quad \sigma_{ij}\boldsymbol{x} = \boldsymbol{x} \quad \text{or} \quad \sigma_{ij}\boldsymbol{x} \gg \boldsymbol{x} \; . \tag{27}$$

So the set of translates $\{\sigma_{ij}\boldsymbol{x} : i, j \in \mathbb{Z}\}$ of a RO state $\boldsymbol{x}$ form a totally ordered set for the partial order on sequences.

Monotonicity of the dynamics implies that weak rotational order is preserved under the dynamics, and strict monotonicity implies that if, at $t = 0$, x is weakly rotationally ordered, then it becomes (and remains) strictly rotationally ordered for positive t. In particular, WRO equilibrium states are in fact RO.

The gradient system (18) induces a well defined semi-flow ϕ_t, $t \geq 0$ at least in the following spaces of configurations of infinite chains:

(i) spatially periodic chains of *type* $(p, q) \in \mathbb{Z} \times \mathbb{N}$: $x_{i+q} = x_i + p \; \forall i \in \mathbb{Z}$;
(ii) chains with bounded width and product topology (the topology of pointwise convergence), metrized for example by $d(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i \in \mathbb{Z}} 2^{-|i|} |x_i - y_i| / (1 + |x_i - y_i|)$.

The existence of the semi-flow arises because the vector field $-\nabla H$ is continuously differentiable, and modulo the translation σ_{01} the spaces are compact.

3.1 Existence of RO Equilibria for Each Mean Spacing

Monotonicity of the gradient flow of (18) can be used to prove the existence of RO equilibria of *all* mean spacings [1, 13], so providing an alternative proof of one of the results of the Aubry-Mather theory [2, 19]. To make the presentation simple we will only prove here the existence of RO spatially periodic equilibrium states, which have rational mean spacing.

Proof. Let $X_{pq} := \{\boldsymbol{x} \in \mathbb{R}^{\mathbb{Z}} : x_{i+q} = x_i + p \quad \forall i \in \mathbb{Z}\}$, the class of type (p, q) spatially periodic sequences. These have mean spacing p/q. The space X_{pq} can be parametrized by the q variables $(x_0, x_1, \ldots, x_{q-1})$. If $\boldsymbol{x} \in X_{pq}$, in system (18) the vector field $(-\frac{\partial H}{\partial x_n}(\boldsymbol{x}))_{n \in \mathbb{Z}}$ is a period-q sequence, hence the space X_{pq} is invariant under the flow.

For sequences $\boldsymbol{x}$ of type (p, q), weak rotational order implies

$$\left\lceil \frac{p}{q}(i-j) \right\rceil \leq x_i - x_j \leq \left\lfloor \frac{p}{q}(i-j) \right\rfloor \tag{28}$$

[4] Golé [13] uses the terminology *cyclically ordered* for rotationally ordered.

(where $\lceil a \rceil$ denotes the smallest integer greater than a and $[a]$ denotes the integral part of a), inequalities that are easily proved by contradiction using the definitions of type (p, q) and WRO.

The quotient space $X^{WRO}_{pq,\sigma} := \{\boldsymbol{x} \in X_{pq}, \boldsymbol{x} \text{ is } WRO\}/\langle\sigma_{01}\rangle$ is forward invariant under the flow and is compact. (To show compactness, without loss of generality take $x_0 \in [0, 1]$, and use (28) for the other x_i's.)

For $\boldsymbol{x} \in X_{pq}$, let $H_{pq}(\boldsymbol{x}) := \sum_{i=0}^{q-1} h(x_i, x_{i+1})$, so H_{pq} is a function of the q variables $(x_0, x_1, \ldots, x_{q-1})$. Now H_{pq} is a continuous function on the compact set $X^{WRO}_{pq,\sigma}$ therefore has a minimum. Also, H_{pq} is a Lyapunov function since

$$\dot{H}_{pq} = \sum_{i=0}^{q-1} \frac{\partial H}{\partial x_i}\dot{x}_i = -\sum_{i=0}^{q-1} \left(\frac{\partial H}{\partial x_i}\right)^2 = -|\nabla H_{pq}|^2 \leq 0 \tag{29}$$

and $\dot{H}_{pq}$ is strictly negative unless $\nabla H_{pq} = \mathbf{0}$, i.e. at equilibrium points. Now if $\boldsymbol{x}$ is WRO then $\phi_t(\boldsymbol{x})$ is RO for all $t > 0$, so there are no equilibria on the boundary of $X^{WRO}_{pq,\sigma}$, and $H_{pq}(\phi_t(\boldsymbol{x})) < H_{pq}(\boldsymbol{x})$ for all $\boldsymbol{x}$ on this boundary. We deduce that the minimum of H_{pq} must be in the interior of $X^{WRO}_{pq,\sigma}$, hence the minimum is a critical point and we have proved the existence of a rotationally ordered equilibrium of type (p, q). □

We shall now investigate the dynamics when there are equilibria.

3.2 Dynamics when There are Equilibria

Here and in the next section, we will extend the discussion to FK models submitted to a "tilted" periodic potential: $h(x, x') = W(x'-x)+V(x)-Fx$, with F constant, i.e. with dynamics subject to a uniform constant additive force (for instance, modelling an electric field on charge density wave materials):

$$\dot{x}_n = W'(x_{n+1} - x_n) - W'(x_n - x_{n-1}) - V'(x_n) + F\ . \tag{30}$$

Again, we will restrict ourselves to spatially periodic solutions. We first investigate what the dynamics are when there are equilibria in $X_{p,q}$. There are RO equilibria when $F = 0$, as shown in the previous section, and if the RO minimum found there is non-degenerate at $F = 0$, there remains an (RO) minimum for $|F|$ small enough.

Suppose there is an equilibrium $\boldsymbol{x}^\star$ in X_{pq} (p, q not necessarily coprime). Let $\boldsymbol{x} \in X_{pq}$. There are integers $j_1 \leq j_2$ such that $\sigma_{0j_1}\boldsymbol{x}^\star \leq \boldsymbol{x} \leq \sigma_{0j_2}\boldsymbol{x}^\star$ and, by order preservation, the same is true for positive times: $\sigma_{0j_1}\boldsymbol{x}^\star \leq \phi_t(\boldsymbol{x}) \leq \sigma_{0j_2}\boldsymbol{x}^\star$, $t > 0$. Therefore the forward orbit of $\boldsymbol{x}$ is stuck in some compact set and hence its omega-limit set $\omega(\boldsymbol{x})$ is not empty. Let $H_{pq}(\boldsymbol{x}) := \sum_{i=0}^{q-1} h(x_i, x_{i+1}) - Fx_i$. We have again $\dot{H}_{pq} = -|\nabla H_{pq}|^2$ and H_{pq} is a Lyapunov function. Then La Salle's Invariance Principle[5] [21] implies that

[5] Salle's Invariance Principle: if V is a Lyapunov function on state space X, i.e. V is non-increasing along orbits in X, then the omega limit set $\omega(\boldsymbol{x})$ of a point $\boldsymbol{x} \in X$ is a set of forward orbits along which V takes some constant value.

$\omega(\boldsymbol{x})$ is a connected set of equilibria with the same value of H_{pq}, $H_{pq} = c$ for some real c. Connectedness is a classical property of omega-limit sets, and monotonicity implies that this set of equilibria has to be totally unordered. □

Generically, $\omega(\boldsymbol{x})$ is a unique equilibrium. Thus a bounded forward orbit of type (p, q) generically converges to an equilibrium, as for parabolic PDEs.

3.3 Dynamics when There are no Equilibria

Consider now the case of "high" tilt. If $|F|$ is large enough, one can show that there are no equilibria of bounded spacing [6]. For instance, for the Standard FK model (20) it is easy to see that this is the case if $|F| > \frac{k}{2\pi}$ since then (in the case $F > 0$), $-\frac{k}{2\pi}\sin 2\pi x_n + F \geq a$ for some positive a and all x_n, and so for an equilibrium we need $(x_n - x_{n-1}) \geq (x_{n+1} - x_n) + a$ for all n, that is an unbounded sequence of spacings.

Say a map $\boldsymbol{u} : \mathbb{R} \to \mathbb{R}^{\mathcal{S}}$ is *periodically sliding* (with period T) if there exists $T > 0$ such that $\boldsymbol{u}(t + T) = \boldsymbol{u}(t) + \mathbf{1}$ for all t, where $\mathbf{1}$ denotes the vector in $\mathbb{R}^{\mathcal{S}}$ with all components equal to 1 (and here $\mathcal{S} = \mathbb{Z}$).

Let's consider again spatially periodic solutions. If there are no spatially periodic equilibria of type (p, q), what happens to the dynamics for this class of initial conditions? We will find that there is a *periodically sliding solution* $\boldsymbol{u}(t + T) = \boldsymbol{u}(t) + \mathbf{1}$ which attracts the orbits of all spatially periodic initial conditions of same mean spacing $\rho = p/q$. All particles of the chain move with the same average velocity $v = 1/T$, which does not depend on the initial condition but depends on ρ and on F.

Theorem 3.1. *[5] If there are no rotationally ordered equilibria in X_{pq} (p, q coprime), then there exists a rotationally ordered periodically sliding solution*

$$\boldsymbol{u}(t + T) = \boldsymbol{u}(t) + \mathbf{1} \qquad \forall\, t \tag{31}$$

in X_{pq}, some $T > 0$, with $\dot{\boldsymbol{u}} \gg \mathbf{0}$ and $\sigma_{mn}\boldsymbol{u} = \boldsymbol{u}((n + mp/q)T)$, which attracts all of $\bigcup_{k\geq 1} X_{kp,kq}$, with phase; i.e. for every $\boldsymbol{x} \in \bigcup_{k\geq 1} X_{kp,kq}$ there exists $\tau \in \mathbb{R}$ such that $||\boldsymbol{x}(t + \tau) - \boldsymbol{u}(t)|| \to 0$ as $t \to \infty$.

Alternatively, (31) can be written $\boldsymbol{u}(t + T) = \sigma_{01}\boldsymbol{u}(t)$, so if we identify $\boldsymbol{u}$ and $\sigma_{01}\boldsymbol{u}$ by taking the quotient space as in Sect. 3.1, the equivalence class $\boldsymbol{u}_\sim(t)$ is a periodic orbit of period T.

Proof. 1. Let

$$X_{pq}^{WRO,+} := \{\boldsymbol{x} \in X_{pq} : \boldsymbol{x} \text{ is WRO and } \dot{\boldsymbol{x}} \geq 0\} \tag{32}$$

the set of weakly rotationally ordered spatially periodic orbits of type (p, q) with non-negative velocity. This set is non-empty because the Aubry-Mather minimal energy equilibrium for $F = 0$ in $X_{p,q}$ is weakly rotationally ordered

and has $\dot{\boldsymbol{x}} = \boldsymbol{F} \geq \boldsymbol{0}$. It is forward invariant by monotonicity. The quotient space $X^{WRO,+}_{pq,\sigma} := X^{WRO,+}_{pq}/\langle\sigma_{01}\rangle$ is compact. Therefore, by Birkhoff's Recurrence Theorem [7][6], $X^{WRO,+}_{pq,\sigma}$ contains a recurrent point. Let $\boldsymbol{u}$ be a representative. Recall that a point $\boldsymbol{x}$ is said to be *recurrent* if it is a limit point of its forward orbit, i.e. there is an sequence of times $t_j \to \infty$ such that $\phi_{t_j}\boldsymbol{x} \to \boldsymbol{x}$ as $j \to \infty$.

2. We now show that if the orbit of $\boldsymbol{u}$ is recurrent in the quotient space $X^{WRO,+}_{pq,\sigma} := X^{WRO,+}_{pq}/\langle\sigma_{01}\rangle$ and is not an equilibrium, then it has to be time-periodic. Suppose that $\boldsymbol{u}$ is not an equilibrium, so that $\dot{\boldsymbol{u}}(0) > \boldsymbol{0}$. Then, by strict monotonicity, $\dot{\boldsymbol{u}}(t) \gg \boldsymbol{0}$ for positive t. Thus $\boldsymbol{u}(t)$ increases and is unbounded, else it would converge to an equilibrium. In fact each $u_i(t)$ is unbounded since RO orbits have to satisfy the constraining inequality (28). Let

$$\begin{cases} \tau_+ := \inf\{\tau \geq 0 : \boldsymbol{u}(\tau) \geq \boldsymbol{u} + \boldsymbol{1}\} \\ \tau_- := \sup\{\tau \geq 0 : \boldsymbol{u}(\tau) \leq \boldsymbol{u} + \boldsymbol{1}\} \,. \end{cases} \tag{33}$$

Then $0 < \tau_- \leq \tau_+ < \infty$. We wish to show that $\tau_- = \tau_+$, so suppose for contradiction that $\tau_- < \tau_+$. Then, by strict monotonicity, given $t_0 > 0$, there exist $\tau'_\pm \in \mathbb{R}$ with $\tau_- < \tau'_- < \tau'_+ < \tau_+$ and

$$\boldsymbol{u}(\tau'_- + t_0) \leq \boldsymbol{u}(t_0) + \boldsymbol{1} \leq \boldsymbol{u}(\tau'_+ + t_0) \,.$$

Note that $\boldsymbol{u}(\tau'_-) \not\leq \boldsymbol{u} + \boldsymbol{1} \not\leq \boldsymbol{u}(\tau'_+)$. There exists a recurrence time $t_1 > t_0$ such that $\boldsymbol{u}(t_1) + \boldsymbol{1}$, $\boldsymbol{u}(\tau'_- + t_1)$, $\boldsymbol{u}(\tau'_+ + t_1)$ are close enough to $\boldsymbol{u} + (N+1)\boldsymbol{1}$, $\boldsymbol{u}(\tau'_-) + N\boldsymbol{1}$, $\boldsymbol{u}(\tau'_+) + N\boldsymbol{1}$ respectively for some positive integer N, that

$$\boldsymbol{u}(\tau'_- + t_1) \not\leq \boldsymbol{u}(t_1) + \boldsymbol{1} \not\leq \boldsymbol{u}(\tau'_+ + t_1) \,. \tag{34}$$

But this contradicts monotonicity from time t_0 to t_1. Hence $\tau_- = \tau_+$. Denoting the common value by T we have $\boldsymbol{u}(T) = \boldsymbol{u}(0) + \boldsymbol{1}$ and hence the desired sliding periodicity.

3. Next we show that $\sigma_{mn}\boldsymbol{u} = \boldsymbol{u}((n + mp/q)T)$. The previous argument applied to $\sigma_{mn}\boldsymbol{u}$ instead of $\boldsymbol{u} + \boldsymbol{1}$, but with $\tau_\pm$ permitted negative since we can extend $\boldsymbol{u}$ to negative times by periodicity, shows that $\sigma_{mn}\boldsymbol{u} = \boldsymbol{u}(\tau)$ for some τ. Thus

$$\boldsymbol{u}((nq + mp)T) = \boldsymbol{u} + (nq + mp)\boldsymbol{1} = \sigma^q_{mn}\boldsymbol{u} = \boldsymbol{u}(q\tau) \,. \tag{35}$$

But $\dot{\boldsymbol{u}} \gg \boldsymbol{0}$, and so $\tau = (n + mp/q)T$.

4. Finally, we show that the orbit of $\boldsymbol{u}$ attracts all of $\bigcup_{k\geq 1} X_{kp,kq}$, with phase. Let $\boldsymbol{x} \in \bigcup_{k\geq 1} X_{kp,kq}$. Then there exists $k \in \mathbb{N}$ such that $\boldsymbol{x} \in X_{kp,kq}$.

[6] Birkhoff's Recurrence Theorem says that every continuous map on a compact metric space has a recurrent point. For a proof not requiring Zorn's lemma, see hint 3.3.4 in [19].

Let

$$\begin{cases} \tau_+(t;\boldsymbol{x}) = \inf\{\tau \in \mathbb{R} : \boldsymbol{u}(\tau + t) \geq \boldsymbol{x}(t)\} \\ \tau_-(t;\boldsymbol{x}) = \sup\{\tau \in \mathbb{R} : \boldsymbol{u}(\tau + t) \leq \boldsymbol{x}(t)\} \ . \end{cases} \tag{36}$$

Then $-\infty < \tau_-(t;\boldsymbol{x}) \leq \tau_+(t;\boldsymbol{x}) < \infty$. If $\tau_-(0;\boldsymbol{x}) = \tau_+(0;\boldsymbol{x}) = \tau$, then $\boldsymbol{x} = \boldsymbol{u}(\tau)$, and hence $\boldsymbol{x}(t) = \boldsymbol{u}(t+\tau)$ for all $t \geq 0$. Else, by strict monotonicity, τ_- is increasing and τ_+ is decreasing. Let $\tau_-^\infty = \sup_{t\geq 0} \tau_-(t;\boldsymbol{x})$ and $\tau_+^\infty = \inf_{t\geq 0} \tau_+(t;\boldsymbol{x})$. Then $\tau_-^\infty \leq \tau_+^\infty$. If $\tau_-^\infty < \tau_+^\infty$, then let $\boldsymbol{y}$ be a limit point of the sequence $(\boldsymbol{x}_k)_{k\in\mathbb{N}} = (\boldsymbol{x}(kT) - k\mathbf{1})_{k\in\mathbb{N}}$ (which exists because the sequence is bounded between $\boldsymbol{u}(\tau_\pm(0;\boldsymbol{x}))$ and $X_{kp,kq}$ is finite-dimensional). It follows that $\tau_\pm(0;\boldsymbol{y}) = \tau_\pm^\infty$. But by strict monotonicity $\tau_-^\infty < \tau_-(t;\boldsymbol{y}) \leq \tau_+(t;\boldsymbol{y}) < \tau_+^\infty$ for $t > 0$. Choose a $t > 0$ and take k large enough that $\boldsymbol{x}_k$ is close enough to $\boldsymbol{y}$ that $\tau_-^\infty < \tau_-(t;\boldsymbol{x}_k) \leq \tau_+(t;\boldsymbol{x}_k) < \tau_+^\infty$. Hence $\tau_-^\infty < \tau_-(t + kT;\boldsymbol{x}) \leq \tau_+(t + kT;\boldsymbol{x}_k) < \tau_+^\infty$ contradicting the definition of $\tau_\pm^\infty$. Thus $\tau_-^\infty = \tau_+^\infty$. Denoting the common value by τ we deduce that $\|\boldsymbol{x}(t + \tau) - \boldsymbol{u}(t)\| \to 0$ as $t \to \infty$. □

Remarks:

(i) Theorem 3.1 implies that if there are no rotationally ordered equilibria of type (p, q), p, q coprime, then there are no equilibria of type (kp, kq) for any $k \geq 1$, else the omega limit set of any type (kp, kq) configuration would be a set of equilibria of the same type, as shown in Sect. 3.2. Hence if there exists an equilibrium of type (kp, kq) for some $k \geq 1$ then there exists a rotationally ordered equilibrium of type (p, q).

(ii) If, instead of thinking of X_{pq} as a set of spatially periodic states of an infinite chain, one thinks of it as a chain of length q with periodic boundary conditions, then one could call these sliding solutions *travelling waves on a ring* (e.g [20])

(iii) Presumably, the "depinning transition" from the case with equilibria to the case with none is generically the bifurcation of a homoclinic orbit to a saddle-node (saddle-node on a cycle bifurcation), which produces a periodic orbit whose frequency (or equivalently, in our context, mean speed) goes like $1/\sqrt{F - F_c}$, where F_c is the transition value of the parameter F.

3.4 Gradient Dynamics of FK Models Under Time-Periodic Forcing

As mentioned earlier, the monotonicity property persists for the gradient dynamics of FK models subject to time-dependent uniform additive or parametric driving forces, where

$$h(x, x') = C(t)W(x' - x) + K(t)V(x) - F(t)x \tag{37}$$

as long as $C(t) \geq 0$ for all t. As a consequence, rotational order and mean spacing are preserved under the dynamics, and the average velocity of the particles does not depend on the initial condition.

In [11] Floría and Mazo investigated the dynamics of the Standard FK model with time-periodic additive forcing

$$\dot{x}_n = x_{n+1} - 2x_n + x_{n-1} - \frac{k}{2\pi} \sin 2\pi x_n + F(t) \tag{38}$$

with $F(t) = \bar{F} + F_{ac} \cos 2\pi\nu_0 t$ (ac-dc forcing). When varying the average $\bar{F}$ of the external driving force, for initial conditions of type (p, q), they observe numerically that the average velocity $\bar{v}$ of the particles is a continuous and non-decreasing function of $\bar{F}$, which is locally constant at rational values of $\bar{v}/\nu_0$. At these rational values, they find stable *resonant* solutions, i.e. solutions satisfying $u_j(t) = u_{j+r}(t - s/\nu_0) + m$ for some integers m, r, s, and for which $\bar{v} = \nu_0(rp + mq)/qs$. That $\bar{v}(\bar{F})$ is continuous and non-decreasing has recently been proved by Hu et al. [17], as well as the stability of the resonant solutions.

Without additive forcing ($F = 0$), a non-zero average velocity of the particles can also be achieved under time-periodic parametric forcing ($C(t) = C(t + \tau)$ or $K(t) = K(t + \tau)$) if the period-1 potential $V(x)$ is not symmetric (e.g. $V(x)$ has one maximum per period at $x = 0$ and one maximum per period at $a \neq 1/2$) – see the chapter by Floría et al. in this book on the ratchet phenomenon.

4 Second Order Local Dynamics

The gradient dynamics case $m = 0$ is an idealised limit, however; in true physical systems there is a small but non-negligible inertial term. Numerical results of Floría, Falo and Mazo suggested that for small positive m, the results presented in the last section for spatially periodic boundary conditions, that is the existence of a globally attracting periodically sliding solution for sufficient tilt F, remained true (though not for large m, as already made clear by [3] and rediscovered by [8]). Many authors had lamented that monotonicity methods did not seem to extend to networks of multicomponent (or equivalently higher order) units, but these numerical results motivated us to look for a partial order for networks of units with second order inertial dynamics, which would be preserved if the inertial terms were small enough compared to the damping. Another motivation for us was a success in the context of strongly damped nonlinear hyperbolic PDEs by Gallay and Raugel [12].

We shall show that under some *overdamping* condition and the same cooperativity condition as before, there is indeed a preserved partial order which makes monotone the inertial dynamics of networks, and this monotonicity is strict if the system is strictly cooperative.

The systems we consider are coupled second order ODEs of the form

$$m_n\ddot{x}_n + \gamma_n\dot{x}_n = f_n(\boldsymbol{x}), \qquad m_n, \gamma_n > 0 \tag{39}$$

for n in some index set $\mathcal{S}$, and C^1 functions f_n satisfying

$$\frac{\partial f_n}{\partial x_j} \geq 0 \quad \forall j \neq n \qquad \text{(cooperativity condition)} \tag{40}$$

and

$$-4m_n\frac{\partial f_n}{\partial x_n} \leq \gamma_n^2 \quad \forall n \qquad \text{(overdamping condition)} \tag{41}$$

in the range of interest. By dividing by γ_n we reduce to the case $\gamma_n = 1$ for all n. By setting $v_n = \dot{x}_n$, we rewrite (39) (with γ_n reduced to 1) as a system of first order ODEs

$$\begin{aligned} \dot{x}_n &= v_n \\ \dot{v}_n &= \frac{1}{m_n}\left(f_n(\boldsymbol{x}) - v_n\right) \end{aligned} \qquad n \in \mathcal{S}\,. \tag{42}$$

We define the *state of a unit* to be the pair $\bar{x}_n = (x_n, v_n) \in \mathbb{R}^2$, and the *state of the network* to be the index set $\bar{\boldsymbol{x}} := (\bar{x}_n)_{n\in\mathcal{S}}$, which we also sometimes write as $\bar{\boldsymbol{x}} = (x_n, \dot{x}_n)_{n\in\mathcal{S}}$. For the state space X we take the set X_0 of all states $\bar{\boldsymbol{x}}$, or various subsets. We assume that conditions hold on X such that (42) defines a local flow ϕ on X.

4.1 Motivation for the Partial Order

To indicate why it is reasonable to expect the flow ϕ of (42) to preserve a partial order for small m_n, consider first the case of a single linear damped oscillator

$$m_n\ddot{x}_n + \dot{x}_n + k_n x_n = 0 \tag{43}$$

(with damping normalised to 1, and $m_n, k_n > 0$). In the overdamped regime $4m_nk_n \leq 1$, the phase portrait (with $v_n = \dot{x}_n$) is shown in Fig. 1. It can be seen that any cone of the form

$$\left\{(x_n, v_n) \in \mathbb{R}^2 : v_n \geq \lambda_n x_n, (2\alpha_n - 1)v_n + \left((1-\alpha_n)\lambda_n^+ - \alpha_n\lambda_n^-\right)x_n \geq 0\right\} \tag{44}$$

with λ_n between the (negative) slopes $\lambda_n^{\pm} = -\frac{1}{2m_n} \pm \sqrt{\frac{1}{4m_n^2} - \frac{k_n}{m_n}}$ of the eigenvectors and $\alpha_n \in [0,1]$, is forward invariant (the cone shown is the case $\lambda_n = -\frac{1}{2m_n}, \alpha_n = \frac{1}{2}$ for which the second condition reduces to $x_n \geq 0$).[7] Turning now to a network of such units with cooperative linear coupling, the coupling makes no difference to $\dot{x}_n$ and adds non-negative contributions to $\dot{v}_n$

[7] A partial order which is preserved by an single overdamped oscillator was in fact found earlier by Levi [22] and Qian et al. [27], independently.

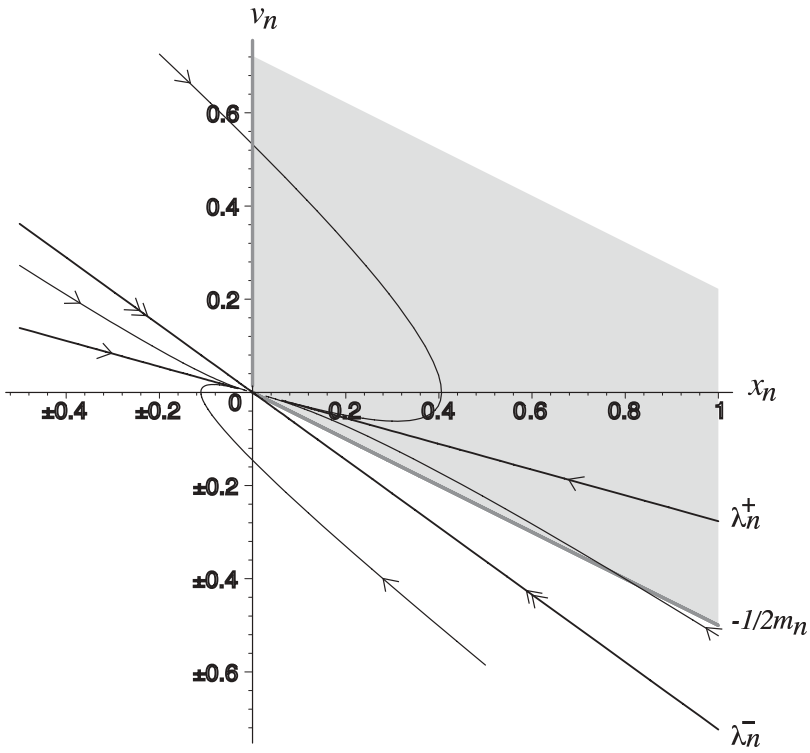


Fig. 1. Phase portrait of (43) for $m_n = 1, k_n = 0.2$. A forward invariant cone is shown in *grey*

from all units with $x_k \geq 0$ because of the cooperativity condition. Thus, the product cone

$$\{(\boldsymbol{x}, \boldsymbol{v}) : v_n \geq \lambda_n x_n, x_n \geq 0 \; \forall n \in \mathcal{S}\} \tag{45}$$

is forward invariant for any choice of λ_n between $\lambda_n^{\pm}$. For simplicity of formulae we will choose $\lambda_n = -\frac{1}{2m_n}, \alpha_n = \frac{1}{2}$ though other choices are possible.

4.2 Order-Preservation

We define a partial order "*less than or equal to*" on X by $\bar{\boldsymbol{x}} \leq \bar{\boldsymbol{y}}$ if for all $n \in \mathcal{S}$

$$\begin{cases} x_n \leq y_n \\ 2m_n \dot{x}_n + x_n \leq 2m_n \dot{y}_n + y_n \; . \end{cases} \tag{46}$$

We use the following standard additional notations

$$\bar{\boldsymbol{x}} \geq \bar{\boldsymbol{y}} \quad \text{if} \quad \bar{\boldsymbol{y}} \leq \bar{\boldsymbol{x}} \qquad \text{"greater than or equal to"} \tag{47}$$

$$\bar{\boldsymbol{x}} < \bar{\boldsymbol{y}} \quad \text{if} \quad \bar{\boldsymbol{x}} \leq \bar{\boldsymbol{y}} \text{ and } \bar{\boldsymbol{x}} \neq \bar{\boldsymbol{y}} \qquad \text{"less than"} \tag{48}$$

$$\bar{\boldsymbol{x}} > \bar{\boldsymbol{y}} \quad \text{if} \quad \bar{\boldsymbol{y}} < \bar{\boldsymbol{x}} \qquad \text{"greater than" .} \tag{49}$$

It will be convenient to define a relation $\ll$ "strictly less than" on X as follows. For $\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}} \in X$, write

$$\bar{\boldsymbol{x}} \ll \bar{\boldsymbol{y}} \quad \text{if} \quad \begin{cases} x_n < y_n \\ 2m_n\dot{x}_n + x_n < 2m_n\dot{y}_n + y_n \end{cases} \quad \text{for all } n \in \mathcal{S}\,. \tag{50}$$

The dynamics of system (42) are said to be *monotone* if given any $\bar{\boldsymbol{x}}(0)$, $\bar{\boldsymbol{y}}(0) \in X$,

$$\bar{\boldsymbol{x}}(0) \leq \bar{\boldsymbol{y}}(0) \implies \bar{\boldsymbol{x}}(t) \leq \bar{\boldsymbol{y}}(t) \tag{51}$$

for all $t > 0$ for which both are defined.
The dynamics are *strictly monotone* if given any $\bar{\boldsymbol{x}}(0)$,$\bar{\boldsymbol{y}}(0) \in X$,

$$\bar{\boldsymbol{x}}(0) < \bar{\boldsymbol{y}}(0) \implies \bar{\boldsymbol{x}}(t) \ll \bar{\boldsymbol{y}}(t) \tag{52}$$

for all $t > 0$ for which both are defined.

For a sequence $\boldsymbol{x} = (x_n)_{n\in\mathcal{S}}$, let

$$\mu(\boldsymbol{x}) := \sup_n \left(-m_n \frac{\partial f_n}{\partial x_n}(\boldsymbol{x}) \right) \tag{53}$$

Proposition 4.1 (Monotonicity of overdamped dynamics).
Suppose that two points $\bar{\boldsymbol{x}}(0)$ and $\bar{\boldsymbol{y}}(0)$ in X satisfy $\bar{\boldsymbol{x}}(0) \leq \bar{\boldsymbol{y}}(0)$, then $\bar{\boldsymbol{x}}(t) \leq \bar{\boldsymbol{y}}(t)$ for $t \geq 0$, as long as (i) both exist and

(ii) $\mu(\bar{\boldsymbol{z}}(t)) \leq \frac{1}{4}$ *for all $\bar{\boldsymbol{z}}(t)$ such that $\bar{\boldsymbol{x}}(t) \leq \bar{\boldsymbol{z}}(t)$ and $\bar{\boldsymbol{z}}(t) \leq \bar{\boldsymbol{y}}(t)$.* (54)

Proof. The monotonicity of the overdamped cooperative dynamics is established in the same lines as for first order dynamics: one first proves monotonicity of the linearised dynamics about a path $\bar{\boldsymbol{x}}(t)$ for which $\mu(\boldsymbol{x}(t)) \leq \frac{1}{4}$ for all $t \geq 0$ (e.g. the m_n are sufficiently small), and then homotopy allows to deduce the same for system (39), provided $\mu \leq \frac{1}{4}$ in the stated region.
(i) The linearisation of (42) about a path $\bar{\boldsymbol{x}}(t)$ is

$$\begin{aligned} \dot{\xi}_n &= \upsilon_n \\ \dot{\upsilon}_n &= \frac{1}{m_n}\left(\sum_k \frac{\partial f_n}{\partial x_k}(\boldsymbol{x})\xi_k - \upsilon_n \right)\,. \end{aligned} \tag{55}$$

Let

$$\begin{aligned} \beta_n &= 2m_n\upsilon_n + \xi_n \\ \gamma_n &= \xi_n e^{t/2m_n} \\ \alpha_n &= \beta_n e^{t/2m_n} \end{aligned} \tag{56}$$

and write $\bar{\boldsymbol{\xi}} = (\bar{\xi}_n)_{n\in\mathcal{S}}$ and $\bar{\boldsymbol{\gamma}} = (\bar{\gamma}_n)_{n\in\mathcal{S}}$ with $\bar{\xi}_n = (\xi_n, \beta_n)$ and $\bar{\gamma}_n = (\gamma_n, \alpha_n)$ respectively. Equation (55) yields

$$\begin{aligned}\dot{\gamma}_n &= \frac{1}{2m_n}\alpha_n \\ \dot{\alpha}_n &= \left(\frac{1}{2m_n} + 2\frac{\partial f_n}{\partial x_n}(\boldsymbol{x}(t))\right)\gamma_n + 2\sum_{k\neq n}\frac{\partial f_n}{\partial x_k}(\boldsymbol{x}(t))\,\gamma_k\end{aligned} \tag{57}$$

or (for short)

$$\dot{\bar{\boldsymbol{\gamma}}} = \bar{B}(\boldsymbol{x}(t))\bar{\boldsymbol{\gamma}}\ .$$

In (57) the term $\frac{1}{2m_n} + 2\frac{\partial f_n}{\partial x_n}(\boldsymbol{x}(t))$ is non-negative by the overdamping condition $\mu(\boldsymbol{x}(t)) \leq \frac{1}{4}$ for all $t \geq 0$, and the partial derivatives of f in the sum are non-negative by the cooperativity condition (40), so the matrix $\bar{B}(\boldsymbol{x}(t))$ is non-negative. It follows that if $\bar{\boldsymbol{\gamma}}(0) \geq \bar{\mathbf{0}}$, that is if $\bar{\boldsymbol{\xi}}(0) \geq \bar{\mathbf{0}}$, then $\bar{\boldsymbol{\gamma}}(t) \geq \bar{\mathbf{0}}$ for all positive t, hence $\bar{\boldsymbol{\xi}}(t) \geq \bar{\mathbf{0}}$ for all positive t, as required.
(ii) Define the homotopy

$$\bar{\boldsymbol{x}}_\lambda(0) = (1-\lambda)\,\bar{\boldsymbol{x}}(0) + \lambda\,\bar{\boldsymbol{y}}(0), \quad \lambda \in [0,1] \tag{58}$$

and let $\bar{\boldsymbol{x}}_\lambda(t)$ be the solution of (42) with initial condition (58), as long as it exists. Let $\bar{\boldsymbol{\xi}}_\lambda(t) = \frac{\partial}{\partial\lambda}\bar{\boldsymbol{x}}_\lambda(t)$. Then

$$\bar{\boldsymbol{\xi}}_\lambda(0) = \bar{\boldsymbol{y}}(0) - \bar{\boldsymbol{x}}(0) \geq \bar{\mathbf{0}}\ . \tag{59}$$

Therefore

$$\bar{\boldsymbol{\xi}}_\lambda(t) \geq \bar{\mathbf{0}} \text{ for } t \geq 0 \text{ as long as } \bar{\boldsymbol{x}}_\lambda(t) \text{ exists and } \mu(\bar{\boldsymbol{x}}_\lambda(t)) \leq \tfrac{1}{4}\ , \tag{60}$$

by part (i). Now

$$\bar{\boldsymbol{x}}_\lambda(t) = \bar{\boldsymbol{x}}(t) + \int_0^\lambda \bar{\boldsymbol{\xi}}_{\lambda'}(t)\,d\lambda' \tag{61}$$

so is trapped between $\bar{\boldsymbol{x}}(t)$ and $\bar{\boldsymbol{y}}(t)$ as long as (60) holds, so the only way for $\bar{\boldsymbol{x}}_\lambda(t)$ to cease to exist or for $\mu(\bar{\boldsymbol{x}}_\lambda(t))$ to exceed $\frac{1}{4}$ is for the hypothesis (54) to fail. In particular

$$\bar{\boldsymbol{y}}(t) = \bar{\boldsymbol{x}}(t) + \int_0^1 \bar{\boldsymbol{\xi}}_\lambda(t)\,d\lambda \geq \bar{\boldsymbol{x}}(t) \tag{62}$$

as long as (54) holds. □

Proposition 4.2 (Strict monotonicity of overdamped dynamics). *If the functions f_n in (39) furthermore satisfy the strict cooperativity condition (13) then $\bar{\boldsymbol{x}}(0) < \bar{\boldsymbol{y}}(0)$ implies that $\bar{\boldsymbol{x}}(t) \ll \bar{\boldsymbol{y}}(t)$ for $t > 0$, as long as (i) both exist and*

(ii) $\mu(\bar{\boldsymbol{z}}(t)) \leq \frac{1}{4}$ *for all* $\bar{\boldsymbol{z}}(t)$ *such that* $\bar{\boldsymbol{x}}(t) \leq \bar{\boldsymbol{z}}(t)$ *and* $\bar{\boldsymbol{z}}(t) \leq \bar{\boldsymbol{y}}(t)$. (63)

Strict monotonicity of the overdamped dynamics in the case of strictly cooperative interactions is proved in a similar (though slightly more technical) way as for the gradient dynamics [6].

5 Overdamped Inertial Dynamics of Frenkel-Kontorova Models

As application of the previous section, we consider again tilted Frenkel-Kontorova chains, for which system (42) takes the form, for all $n \in \mathbb{Z}$

$$\begin{aligned} \dot{x}_n &= v_n \\ \dot{v}_n &= \frac{1}{m}\left(W'(x_{n+1}-x_n) - W'(x_n - x_{n-1}) - V'(x_n) - v_n + F\right) \end{aligned} \tag{64}$$

with $m_n = m > 0$ for all n, F constant and non-negative, and V,$W \in \mathcal{C}^2$ satisfying $V(x+1) = V(x)$ (periodicity), $W''(x) \geq b > 0$ (convexity / cooperativity). Here $\mu(\boldsymbol{x}) = \sup_n m(W''(x_{n+1}-x_n) + W''(x_n - x_{n-1}) + V''(x_n))$. For instance, for the Standard FK model, the second equation in (64) reads

$$\dot{v}_n = \frac{1}{m}\left(x_{n+1} - 2x_n + x_{n-1} + \frac{k}{2\pi}\sin 2\pi x_n - v_n + F\right) \tag{65}$$

and $\mu(\boldsymbol{x}) \leq m(2+k)$, so the overdamping condition (54) is satisfied for all states as soon as

$$0 < m \leq \frac{1}{4(2+k)} \cdot$$

We restrict here to the space X of states with bounded spacings and velocities, $\sup_{n\in\mathbb{Z}} |x_n - x_{n-1}| < \infty$ and $sup_{n\in\mathbb{Z}}|v_n| < \infty$, with metric given by supremum norm, for which system (64) defines a local flow ϕ. This local flow commutes with the group of *translations* $\{\bar{\sigma}_{i,j} : (i,j) \in \mathbb{Z}^2\}$ defined by

$$(\bar{\sigma}_{i,j}\bar{\boldsymbol{x}})_n = (x_{n+i} + j, v_{n+i}),\ \forall n \in \mathbb{Z}\ . \tag{66}$$

The *width* of a state is defined as for the gradient dynamics (24) (with $\bar{\sigma}_{i,j}$ instead of $\sigma_{i,j}$ and states $\bar{\boldsymbol{x}}$ instead of configurations $\boldsymbol{x}$), and is preserved under the forward dynamics. Configurations with bounded width have well-defined mean spacing, which is also preserved under the forward dynamics.

Say $\bar{\boldsymbol{x}}$ is *weakly rotationally-ordered* (WRO) if for all $i,j \in \mathbb{Z}$ either

$$\bar{\sigma}_{i,j}\bar{\boldsymbol{x}} \leq \bar{\boldsymbol{x}} \quad \text{or} \quad \bar{\sigma}_{i,j}\bar{\boldsymbol{x}} \geq \bar{\boldsymbol{x}}\ . \tag{67}$$

(Equivalently, $\bar{\boldsymbol{x}}$ has width ≤ 1.) Similarly, say $\bar{\boldsymbol{x}}$ is *(strictly) rotationally-ordered* if for all $i,j \in \mathbb{Z}$ either

$$\bar{\sigma}_{i,j}\bar{\boldsymbol{x}} \ll \bar{\boldsymbol{x}} \quad \text{or} \quad \bar{\sigma}_{i,j}\bar{\boldsymbol{x}} = \bar{\boldsymbol{x}} \quad \text{or} \quad \bar{\sigma}_{i,j}\bar{\boldsymbol{x}} \gg \bar{\boldsymbol{x}}\ . \tag{68}$$

By strict monotonicity, if $\bar{\boldsymbol{x}}(0)$ is weakly rotationally-ordered then $\bar{\boldsymbol{x}}(t))$ is rotationally ordered for $t > 0$.

Global Semi-Flow

For $N \in \mathbb{N}$ define X_N to be the set of states in X such that

$$|x_n - x_{n-1}| \leq N \tag{69}$$
$$|2m(v_n - v_{n-1}) + (x_n - x_{n-1})| \leq N \tag{70}$$

and let

$$\mu_N = m \ \sup\{W''(x_{n+1} - x_n) + W''(x_n - x_{n-1}) + V''(x_n)\} \tag{71}$$

where the supremum is taken over the x_{n-1}, x_n, x_{n+1} such that

$$|x_n - x_{n-1}| \leq N \quad \text{and} \quad |x_{n+1} - x_n| \leq N \ .$$

Note that every $\bar{\boldsymbol{x}} \in X$ is in some X_N because there exist L_1, L_2, V_1, V_2 such that for all $n \in \mathbb{Z}$, $(x_n - x_{n-1}) \in [L_1, L_2]$ and $v_n \in [V_1, V_2]$; then take N to be the integer above $2m(V_2 - V_1) + \max\{|L_1|, |L_2|\}$. System (64) defines a global semi-flow on X_N for $\mu_N \leq \frac{1}{4}$. This is because $\bar{\boldsymbol{x}} \in X_N$ is equivalent to

$$\bar{\sigma}_{-1,-N}\, \bar{\boldsymbol{x}} \leq \bar{\boldsymbol{x}} \leq \bar{\sigma}_{-1,N}\, \bar{\boldsymbol{x}} \ . \tag{72}$$

Thus for $\mu_N \leq \frac{1}{4}$ Proposition 4.1 implies that X_N is preserved under the forward flow. The vector field is Lipschitz on X_N (since V and W are $\mathcal{C}^2$ and V is periodic), hence the flow is defined for all positive times.

Dynamics of Spatially Periodic States

We call *spatially periodic states of type* (p, q), the states in the set

$$\bar{X}_{p,q} := \{\bar{\boldsymbol{x}} \in X : x_{n+q} = x_n + p, v_{n+q} = v_n \ \forall n \in \mathbb{Z}\} \ ;$$

this set is invariant under the dynamics of (64).

We now show that the results on the gradient dynamics of spatially periodic configurations described in Sect. 3 extend to the overdamped inertial dynamics of spatially periodic states.

Dynamics When There are no Equilibria

We say that $\bar{\boldsymbol{x}}(t)$ is *periodically sliding* with *period* T if

$$\bar{\boldsymbol{x}}(t + T) = \bar{\boldsymbol{x}}(t) + \bar{\mathbf{1}} \qquad \forall t$$

where $\bar{\mathbf{1}}$ denotes the state with $(\bar{\mathbf{1}})_n$. In absence of equilibria, we find again that there exists a unique periodically sliding state which attracts all spatially periodic solutions of same mean spacing.

Theorem 5.1 (Periodically sliding solutions). *If there are no rotationally ordered equilibria in $\bar{X}_{p,q}$ then there exists a rotationally ordered periodically sliding solution $\bar{\boldsymbol{u}}(t+T) = \bar{\boldsymbol{u}}(t) + \bar{\mathbf{1}}$ with $T > 0$ in $\bar{X}_{p,q}$, with $\dot{\bar{\boldsymbol{u}}} \gg \bar{\mathbf{0}}$ and $\bar{\sigma}_{m,n}\bar{\boldsymbol{u}} = \bar{\boldsymbol{u}}((n + m\frac{p}{q})T)$, which attracts all of $\cup_{k\geq 1}\bar{X}_{kp,kq}$ with phase, i.e. for every $\bar{\boldsymbol{x}} \in \cup_{k\geq 1}\bar{X}_{kp,kq}$ there exists $\tau \in \mathbb{R}$ such that $\|\bar{\boldsymbol{x}}(t+\tau) - \bar{\boldsymbol{u}}(t)\| \to 0$ as $t \to \infty$.*

The proof follows the lines of that of Theorem 3.1 (see [6]).

The question of existence of periodically sliding solutions in the classes of spatially periodic states of type (p, q) has recently been addressed by Katriel in [20] for inertial but *not necessarily overdamped* dynamics of tilted FK models. There, no monotonicity property is assumed and the proof consists of formulating this existence problem as a fixed point problem in a Banach space and then uses the Schauder Fixed Point Theorem. Remain open the interesting questions of uniqueness and stability of those solutions.

Dynamics When There are Equilibria

In presence of equilibria, we find again that orbits converge to (sets of) equilibria. More precisely, suppose that there exist equilibria in $\cup_{k\geq 1}\bar{X}_{kp,kq}$. Note from Theorem 5.1 that in this case there must be a rotationally ordered equilibrium in $\bar{X}_{p,q}$.

Theorem 5.2. *If there exists a rotationally-ordered equilibrium in $\bar{X}_{p,q}$ then for all $\bar{\boldsymbol{x}} \in \cup_{k\geq 1}\bar{X}_{p,q}$ the ω-limit set $\omega(\bar{\boldsymbol{x}})$ is an unordered connected subset of the set of equilibria in $\bar{X}_{kp,kq}$, some k, with constant $H_{kp,kq}$.*

Generically $\omega(\bar{\boldsymbol{x}})$ is a single equilibrium.

Proof. The initial condition $\bar{\boldsymbol{x}}$ is bounded by certain translates of the equilibrium. On $\bar{X}_{p,q}$, define the total energy of a one-period segment of the chain

$$H_{p.q}(\bar{\boldsymbol{x}}) = \sum_{n=0}^{q-1} W(x_{n+1} - x_n) + V(x_n) - Fx_n + \tfrac{1}{2}m\dot{x}_n^2 \,. \tag{73}$$

Now $H_{p.q}$ is a Lyapunov function since $\frac{dH_{pq}}{dt}(\bar{\boldsymbol{x}}) = -\sum_{n=0}^{q-1} v_n^2 \leq 0$, and $\dot{H}_{pq}(\bar{\boldsymbol{x}})$ is strictly negative unless $\bar{\boldsymbol{x}}$ is an equilibrium. Thus it follows from La Salle's invariance principle [21] that there exists $E \in \mathbb{R}$ such that $\omega(\bar{\boldsymbol{x}})$ is a connected subset of the set of equilibria in $\bar{X}_{kp,kq}$ with $H_{kp,kq} = E$. That $\omega(\bar{\boldsymbol{x}})$ is unordered follows from [28]. □

For the overdamped inertial dynamics of the Standard FK model with additive driving force $F(t) = \bar{F} + F_{ac}\cos 2\pi\nu_0 t$, Hu et al. [17] have obtained the same results as those mentioned in Sect. 3.4 for the gradient dynamics.

6 Further Discussion and Open Problems

1. The preserved partial order of Sect. 4 can be extended to networks of more general type than second order mechanical units; e.g. the units can be any dimensional systems of first order ODEs such that the linearized forward dynamics preserves some cone and the coupling respects the product cone.
2. Theorem 3.1 and 5.1 are also valid for finite FK chains [5]. In the gradient dynamics case, elements of a proof were first presented by Middleton [24] in the case of finite chains and then by Floría and Mazo [11] for infinite chains, but their proofs were incomplete.
3. In our presentation of gradient and overdamped dynamics of Frenkel-Kontorova models with constant uniform driving force, we have restricted ourselves to classes of spatially periodic configurations of type (p, q). Can Theorems 1 and 2 be extended to genuinely infinite dimensional classes of initial conditions? The issue is that strict monotonicity is not enough. The next two items elaborate on this question for two particular cases.
4. What is the fate of spatially non-periodic configurations of rational mean spacing e.g. discommensurations, i.e. $u_n \to v_n$ as $n \to \infty$ and $\to v_{n+j} - i$ as $n \to -\infty$ for some $\boldsymbol{v}$ of type (p, q) with $iq - jp = \pm 1$ (these are called discommensurations of type $(p, q)\pm$). For instance
   a) if there are no type (p, q) equilibria does every configuration of mean spacing p/q and bounded spacing converge pointwise to the periodically sliding type (p, q) solution, with phase?
   b) if there are type (p, q) equilibria but no equilibrium discommensurations of type $(p, q)+$, is there a periodically sliding type $(p, q)+$ discommensuration?[8] What is its basin of attraction?

   These questions are open in both the gradient case and the overdamped case.
5. It should be possible to extend Theorems 3.1 and 5.1 to weakly rotationally ordered configurations with irrational mean spacing. An attempt was presented in [5] but an error was pointed out in [6]. Assuming that a RO sliding solution with irrational mean spacing exists, what would be its basin of attraction?
6. For FK models with fixed constant tilt F, can one show that the sliding velocity $v(\rho)$ (defined to be $1/T$, with T the sliding period) is a continuous function of the mean spacing at irrationals (assuming the solution to question 5 is found) and $\lim_{\rho \nearrow \frac{p}{q}\pm} v(\rho)$ exist and equal $v(\frac{p}{q}\pm)$ (assuming solution to question 4b) and $v(\frac{p}{q}\pm) \geq v(\frac{p}{q})$ with generic strict inequality if $v(\frac{p}{q}\pm) \neq 0$? (as suggested by the numerics of Floría and Mazo[11]). This result, namely the continuous dependence of the sliding velocity on

[8] In the case of *integer* mean spacing, existence of a travelling discommensuration for gradient dynamics can be inferred from [10] and was obtained independently in [9].

the mean spacing, has been obtained for discrete time FK models, see the chapter by R. Coutinho and B. Fernandez in this book.

References

1. S. Angenent, *The periodic points of an area-preserving twist map*, Commun. Math. Phys **115**, 353–74 (1988). 247, 248
2. S. Aubry and P.-Y. Le Daeron, *The discrete Frenkel-Kontorova model and its extensions*, Physica D **8**, 381–422 (1983). 248
3. S. Aubry and L. de Seze, *Dynamics of a charge-density wave in a lattice*, in: Festkörperprobleme (Advances in Solid State Physics), **XXV**, 59–69, P. Grosse (ed), Vieweg, Braunschweig (1985). 253
4. C. Baesens, Variational Methods in Solid-State Physics, Lecture Notes, University of Warwick (1994).
5. C. Baesens and R.S. MacKay, *Gradient dynamics of tilted Frenkel-Kontorova models*, Nonlinearity **11**, 949–964 (1998). 247, 250, 261
6. C. Baesens and R.S. MacKay, *A novel preserved partial order for cooperative networks of units with overdamped second order dynamics, and application to tilted Frenkel-Kontorova chains*, Nonlinearity **17**, 567–580 (2004). 250, 257, 260, 261
7. G.D. Birkhoff, *Dynamical Systems*, (Providence, RI: American Mathematical Society, 1927). 251
8. A. Carpio and L.L. Bonilla, *Oscillatory wave fronts in chains of coupled nonlinear oscillators*, Physical Review E(3) **67** (2003), no. 5, 056621, 11 p. 246, 253
9. A. Carpio, S.J. Chapman, S. Hastings and J.B. McLeod, *Wave solutions in a discrete reaction-diffusion equation*, Euro. Jnl Appl. Math. **11**, 399–412 (2000). 261
10. S.-N. Chow, J. Mallet-Paret and W. Shen, *Traveling waves in lattice dynamical systems*, J. Diff. Eqns **149**, 248–291 (1998). 261
11. L.M. Floría, J.J. Mazo, *Dissipative Dynamics of the Frenkel-Kontorova model*, Advances in Physics **45**(6), 505–598 (1996). 246, 253, 261
12. T. Gallay and G. Raugel, *Stability of travelling waves for a damped hyperbolic equation*, Z Angew M P **48**, 451–479 (1997). 253
13. C. Golé: *Symplectic Twist Maps: Global Variational Techniques, Advanced Series in Nonlinear Dynamics* **18**, World Scientific (2001), and references therein. 248
14. M. Hirsch, *Systems of differential equations which are competitive or cooperative 1: Limit sets*, SIAM J. Appl. Math. **13** 167–179 (1983). 241
15. M. Hirsch, *Systems of differential equations which are competitive or cooperative II: convergence almost everywhere*, SIAM J. Math. Anal.**16**, 423–39 (1985). 241
16. M. Hirsch, *Stability and convergence in strongly monotone dynamical systems*, J. reine angew. Math. **383** 1–53 (1988). 245
17. B. Hu, W-X Qin and Z. Zheng, *Rotation number of the overdamped Frenkel-Kontorova Model with AC-driving*, preprint 2004. 253, 260
18. E. Kamke, *Zur Theorie der Systeme Gewoknlicher Differentialgliechungen*, II, Acta Math. **58**, 57–85 (1932). 241
19. A. Katok and B. Hasselblatt: Introduction to the Modern Theory of Dynamical Systems, (Cambridge, CUP, 1995). 248, 251
20. G. Katriel, *Existence of travelling waves in discrete sine-Gordon rings*, preprint 2004. 252, 260

21. J.P. La Salle, *An invariance principle in the theory of stability*, in Int. Symp. on Differential Equations and Dynamical Systems ed J Hale and J P La Salle (London: Academic), p 277 (1967). 249, 260
22. M. Levi, *Nonchaotic behavior in the Josephson Junction*, Phys. Rev. A bf 37(3) 927–931 (1988). 254
23. H. Matano, *Strongly order-preserving local semi-dynamical systems – Theory and applications*, in Res. Notes in Math., 141, Semigroups, Theory and Applications, eds H. Bresis, M.G. Crandall and F. Kappel, eds, vol. 1, (Longman Scientific and Technical, London) pp 178–185 (1986). 245
24. A.A. Middleton, *Asymptotic uniqueness of the sliding state for charge density waves*, Phys Rev. Lett. **68**, 670–673 (1992). 261
25. M. Müller, *Uber das Fundamenthaltheorem in der Theorie der gewohnlichen Differentialgleichungen*, Math. Zeit. **26** 619–645 (1926). 241
26. M.H. Protter and H.F. Weinberger: Maximum Principles in Differential Equations, (Springer-Verlag, New-York, 1984). 241
27. M. Qian, W Shen and J. Zhang, *Global behavior in the dynamical equation of J-J type*, J. Diff. Eqns **71**(2), 315–333 (1988). 254
28. H.L. Smith: Monotone Dynamical Systems, Providence, AMS Mathematical surveys and monographs **41** (1995). 241, 245, 260

The Frenkel-Kontorova Model

L.M. Floría[1,3], C. Baesens[2] and J. Gómez-Gardeñes[1,3]

[1] Dpt. of Theory and Simulation of Complex Systems, ICMA, CSIC-Universidad de Zaragoza, E-50009 Zaragoza, Spain
[2] Mathematics Institute, University of Warwick. Coventry CV4 7AL, UK
[3] Instituto de Biocomputación y Física de Sistemas Complejos, and Dpto. de Física de la Materia Condensada, Universidad de Zaragoza, E-50009 Zaragoza, Spain
claude@maths.warwick.ac.uk, floria@posta.unizar.es, gardenes@unizar.es

1 Introduction – Presentation of the Model

In the preface to his monograph on the structure of Evolutionary Theory [1], the late professor Stephen Jay Gould attributes to the philosopher Immanuel Kant the following aphorism in Science Philosophy: "Percepts without concepts are blind; concepts without percepts are empty". Using with a bit of freedom these Kantian terms, one would say that a scientific model is a framework (or network) of interrelated concepts and percepts where experts build up scientific consistent explanations of a given set of observations. Good models are those which are both, conceptually *simple* and *universal* in their perceptions. Let us illustrate with examples the meaning of this statement.

The mathematical pendulum, i.e. a Newtonian mass point of mass m and (angular) position u, subject to a sinusoidal potential $V(u)$, with

$$\text{energy} = \frac{1}{2}m\dot{u}^2 + V(u) \tag{1}$$

is an archetype of good model. The variety of its applications makes it ubiquitous in Physics. Figure 1 shows the formal equivalence of a mathematical pendulum and the Josephson effect between two superconducting electrodes separated by a thin layer of an insulating material. Here the angular variable φ is the so-called gauge invariant superconducting phase difference

$$\varphi = \theta_1 - \theta_2 - (2\pi/\Phi_0)\int_1^2 A dl \tag{2}$$

where A is the electromagnetic vector potential, $\theta_{1(2)}$ are the phases of the wave function of the superconducting state at each electrode, and Φ_0 is a constant called the flux quantum unit.

L.M. Floría, C. Baesens and J. Gómez-Gardeñes: *The Frenkel-Kontorova Model*, Lect. Notes Phys. **671**, 209–240 (2005)
www.springerlink.com

A second good example is the Ising model, a familiar model in modern Statistical Mechanics. Approaches based on this model play there a role as basic as pendulum in General Physics does. The Ising model is a lattice or network of N interconnected two-state ($s_j = \pm 1$) nodal elements interacting with energy $J_{ij}s_is_j$ per link, and N is macroscopically large. The complex singular macroscopic behaviour observed in the so-called critical phenomena, has in this conceptual-perceptual framework a referent of rigorous basic understanding. The simplicity of the two-state (on-off, up-down) nodal element on more or less complex connectivity patterns, pervades current theoretical studies in neural networks. It also permeates recent approaches to biological, social and communication networking studies, where interesting generalizations of the model are often needed. In particular, more general dynamical variables at nodes are clearly required in many interesting issues concerning these applications.

The model of interest here belongs, as the pendulum and Ising model, to that class of scientific models which along the years have provided insightful conceptual tools and perceptions in the rigorous analysis of a wide range of theoretical problems. We refer to this model as the Frenkel-Kontorova (FK) model, though other names are sometimes used in the vast literature on this model. Nowadays, in Statistical and Nonlinear Physics it plays a basic role with important applications in materials and condensed matter systems as well as in several micro- and nano-scale technologies [2, 3].

The *Standard* FK model can be seen as a one-dimensional lattice of identical pendula oscillating in parallel planes, coupled to nearest neighbors by identical (linear) torsion springs.

Alternatively, viewed as a model for spatially modulated structures in solid state physics, the (more general) FK model is a chain of Newtonian particles in one dimension, connected by springs and placed in a spatially periodic potential which represents, for instance, the interaction with a substrate. The total potential energy of the system is formally[1]

$$H(\{u_i\})) = \sum_i V(u_i) + W(u_i, u_{i+1}) \tag{3}$$

where u_i denotes the position of particle i, V is the (on-site) external potential, and W is the interaction potential between (nearest) neighbors. We shall denote by $h(u_i, u_{i+1})$ the local potential energy at site i:

$$h(u_i, u_{i+1}) = V(u_i) + W(u_i, u_{i+1}) \; . \tag{4}$$

In the Standard FK model, u_i denotes the (one-dimensional) angle variable at site i, and the potential functions have the following forms: $V(u) = \frac{K}{4\pi^2}(1 - \cos 2\pi u)$, and $W(u, u') = \frac{1}{2}(u' - u)^2 - \mu(u' - u)$, where μ denotes the natural length of the springs. More generally, we call here (unforced) *Frenkel-Kontorova models*, the systems for which the local potential V is periodic:

[1] *Formally*, because the sum in (3) typically diverges in the case of an infinite chain.

$V(u+1) = V(u)$, to fix ideas, and the interaction function $W(u, u')$ depends only on the difference $u' - u$, i.e., (abusing notations) $W(u, u') = W(u' - u)$ or $W(\Delta u)$ (using the notation $\Delta u = u' - u$), and furthermore W is a (strictly) convex function, so that $W''(u) > 0$.[2] This convexity property of the interactions turns out to be a crucial ingredient both for the well established (rigorous) theory of equilibria in the thermodynamic (infinite system size) limit, that we will summarize in Sect. 2, and for the dissipative dynamics that we will develop in further sections. Much less general theory is available for the cases in which $W(\Delta u)$ is not a convex function.

Motivated in the 30's and 40's (XXth century) by studies on plasticity of solids i.e. dynamics and thermodynamics of localized lattice imperfections, the model of coupled pendula or other nonlinear oscillators serves today as the referent name of a streamline in the little history of solid state physics research. Localized elementary excitations under terms as diverse as lattice (or discrete) skyrmions, vortices, fronts, solitons, kinks and breathers have emerged as the most convenient basic descriptors of the ubiquitous complex phenomena associated with the competition of

- self-focussing effect of the nonlinear (non-quadratic) on-site potential $V(u)$, and
- dispersive effect of coupling between sites.

These are the essential ingredients captured in a simplest (though non trivial) way by the FK model. A good indication of the offsprings vigor of this conceptual-perceptual framework is the recent interest (see [4]), both theoretical and experimental, in the issues of localization and transport in Nonlinear lattices. Among the ample variety of current experimental lines of research with which the model is concerned, we now briefly refer to two motivating examples. The first one comes from superconducting technologies[3] at the micro-scale (micrometer sized devices) which are based on the above mentioned Josephson effect. The second example deals with the elastic properties at the nanoscale level of the DNA double strand.

1.1 A JJ Parallel Array

The superconducting circuit sketched in Fig. 1.e [5] consists of two parallel superconducting wires with equispaced Josephson junctions (JJ) in between, which are labeled by j. The relevant nodal variables φ_j are the gauge invariant phase differences across the junctions. The equation for their time evolution is

$$\ddot{\varphi}_j + \Gamma\dot{\varphi}_j + \sin\varphi_j = \lambda(\varphi_{j+1} - 2\varphi_j - \varphi_{j-1}) + i_{ext} \tag{5}$$

[2] We also assume that V and W are twice differentiable.

[3] These technologies have practical uses in quantum metrology, radio-frequency emission, magnetic field and photo-detection, as well as current attempts to (solid state based) quantum computation devices.

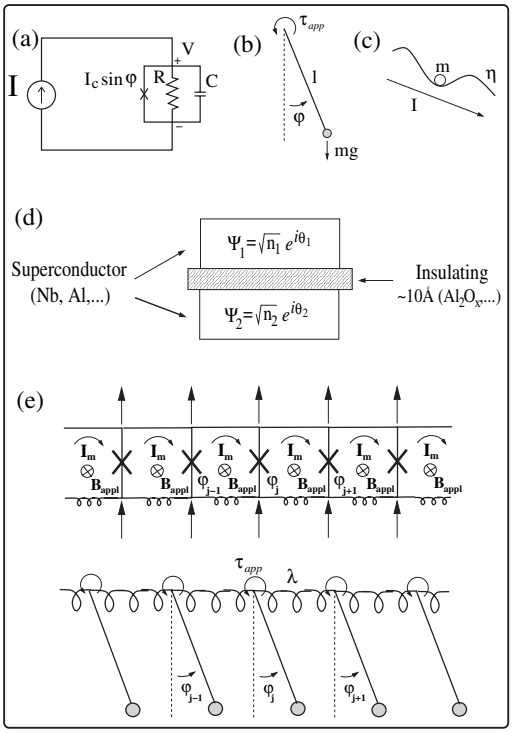


Fig. 1. Equivalence of Josephson junction arrays and chains of coupled pendula. The Josephson current and voltage between the two superconductors of the Josephson junction represented in (**d**) are functions of the phase difference φ introduced in (2): $I_J = I_c \sin\varphi$ and $V = \hbar/2e\; d\varphi/dt$. The RCSJ model sketched in (**a**) describes all the contributions for the current in the Josephson junction: $I = I_J + I_R + I_C$. Using the φ-dependence of the voltage and the Josephson current and after normalizing, the equation for the currents becomes $i_{ext} = \ddot{\varphi} + \Gamma\dot{\varphi} + \sin\varphi$. This equation has the same functional form as the one describing the motion of a damped pendulum with torque (**b**) or of a particle in a tilted washboard potential (**c**). The expression of the single Josephson junction is used for the description of the Josephson junction parallel array (**e**) to obtain (5) which, in this case, is formally equivalent to the equation describing the dynamics of a chain of damped pendula with torque (τ_{app}) coupled by torsion springs [5]

where the parameter λ measures the importance of the induced fields, the damping parameter Γ that of the resistive normal current, and i_{ext} the bias external current. The presence of a magnetic field is felt by the electromagnetic circuit as an imposed average torsion per unit length on each field φ_j. Denoting by f_0 the magnetic flux (in quantum flux units) through a single plaquette, the boundary conditions are given by $\varphi_0 = \varphi_1 - 2\pi f_0$ and $\varphi_{N+1} = \varphi_N + 2\pi f_0$ for the case of open-ended arrays, and $\varphi_{N+1} = \varphi_1 - 2\pi n_k$ and $\varphi_0 = \varphi_N - 2\pi n_k$ for ring arrays. The integer n_k is the number of fluxons trapped in the array.

Equation (5) is easily recognized as describing the dynamics of a 1d chain of damped pendula with torque, coupled by torsion springs. Experiments on parallel JJ arrays and their consistent interpretation based on the dynamics of the FK model were reported in [6]. More recently, several designed (periodically inhomogeneous) parallel JJ arrays were experimentally studied showing ratchet transport of fluxon matter, as theoretically predicted [7]. Using the basic tools and concepts from the rigorous theory of FK models, we will show in Sect. 5 that some classes of parametrically driven FK models exhibit collective (non-thermal) ratchet transport.

1.2 Unzipping DNA

Replication and transcription are two basic functions performed by the DNA molecule both inside living cells and in laboratory routine essays. These functions require the opening of base pairs (A-T, G-C, ...) as schematically illustrated in Fig. 2. Motivated by the possible role of bubble opening in thermal

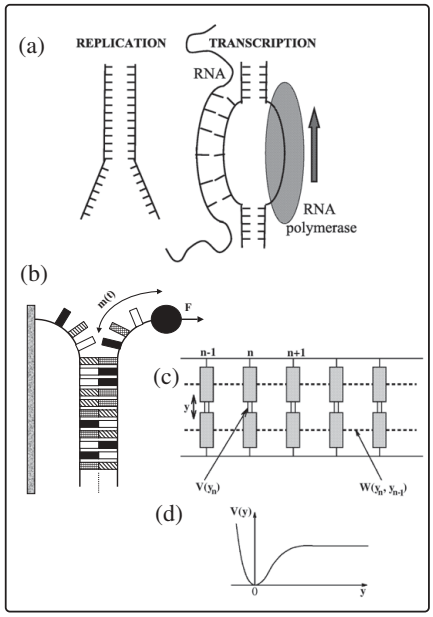


Fig. 2. (**a**) Schematic representation of the Replication and Transcription of the DNA. The DNA unzipping involved in the two processes is described using a constant force scheme based on the Peyrard-Bishop-Dauxois sequence dependent model, as sketched in (**b**). In the PBD model, sketched in (**c**), the relevant on-site variable is the distance y_j between the two bases of the pair j. The potential $V(y_j)$ accounts for the interaction between these two nucleotides (A-T or G-C), and $W(y_j, y_{j+1})$ is the energy of interaction between neighboring pairs j and $j+1$. (**d**) Shape of the potential $V(y)$ given by (6)

denaturation of double stranded DNA chains, the model of Peyrard-Bishop-Dauxois, PBD for short, tries to capture in a workable simple way the local energy balance which governs the physics of base pair opening.

The relevant variable y_j at pair j is the inter-base distance, so that pair opening means "big y". Following the intermolecular potential theory, the following functional form for the on-site potential energy has the appropriate shape:

$$V(y) = D\left(e^{-ay} - 1\right)^2 \tag{6}$$

while the coupling through the double strand of sugar backbone can be assumed either quadratic, as in the original version [8], or sensibly corrected as in [9]:

$$W(y_j, y_{j+1}) = \frac{1}{2}C(y_{j+1} - y_j)^2\left(1 + \rho e^{-\alpha(y_{j+1}+y_j)}\right). \tag{7}$$

The PBD model is a generalized[4] FK model for the elasto-plastic properties of DNA base pair opening. Interesting experiments on DNA unzipping have been already reported [10]. One should note that there is no periodicity symmetry in real DNA of living cells, though periodic blocks in the "junk" chromosomal DNA are well known to be overabundant [11]. Thus the lack of homogeneity is essential for the basic description of many observed phenomena in living double stranded DNA molecules. We stress this point here, as an example of the added complexity in this particular use of the FK model in DNA prospective technologies.

2 Equilibrium States

In condensed matter, a physical motivation for studying equilibrium states of the FK model came from the observed abundance of modulated (structured) phases in minerals as well as in man-made materials and/or compounds, and the need to understand the peculiar multiphase diagrams shown by experiments.

We shall consider here the thermodynamic limit, so equilibrium configurations (or states) will be represented by doubly infinite sequences (u_i), $-\infty < i < +\infty$. Balance of forces at each site n yields[5]

$$0 = -\frac{\partial H}{\partial u_n} = -\frac{\partial(h(u_{n-1}, u_n) + h(u_n, u_{n+1}))}{\partial u_n} \tag{8}$$

$$= W'(u_{n+1} - u_n) - W'(u_n - u_{n-1}) - V'(u_n) \tag{9}$$

for the (general) FK model, and more specifically for the Standard FK model

$$u_{n+1} - 2u_n + u_{n-1} - \frac{K}{2\pi}\sin 2\pi u_n = 0\,. \tag{10}$$

[4] *generalized*, in the sense that the on-site potential $V(y)$ is not periodic.

[5] Though the energy H in (3) may be infinite, its gradient is well defined.

There is a one-to-one correspondence between equilibrium configurations (u_n) of an FK model and orbits $((u_n, p_n))$ of a symplectic twist map of the cylinder, which can be obtained from (9) using a Legendre transform, as usually done to go from the Lagrangian formulation of a problem to the Hamiltonian one. Setting $p_n = -\frac{\partial h(u_n,u_{n+1})}{\partial u_n} = W'(u_{n+1}-u_n)-V'(u_n)$, the convexity condition $W'' > 0$ allows to invert this expression to obtain u_{n+1} as a function of u_n and p_n, and then we use $p_{n+1} = W'(u_{n+1} - u_n)$. For instance, for the Standard FK model, we easily obtain from (10):

$$u_{n+1} = u_n + p_n + \frac{K}{2\pi}\sin 2\pi u_n \tag{11}$$

$$p_{n+1} = p_n + \frac{K}{2\pi}\sin 2\pi u_n \ . \tag{12}$$

This is the familiar Standard map, paradigm example of a twist map[6] of the cylinder.

Of particular importance are the *minimum energy configurations*, m.e.c. for short, which are sequences (u_j) such that, given any integers $m < n$,

$$H_{mn} \equiv \sum_{j=m}^{n-1} h(u_j, u_{j+1}) \tag{13}$$

is (globally) minimal with respect to all variations of $(u_{m+1}, \ldots, u_{n-1})$, keeping u_m and u_n fixed. They correspond to action-minimizing orbits of the associated twist map.

M.e.c. are also local minima of the energy (3), meaning equilibrium configurations for which the quadratic form in the series expansion of the energy around the equilibrium is positive or zero:

$$\delta H \simeq \frac{1}{2}\sum_{m,n} \frac{\partial^2 H}{\partial u_n \partial u_m}\delta_n\delta_m \geq 0 \tag{14}$$

(for square summable (δ_n)). Local minima are (linearly) stable equilibria. For stable equilibria, the eigenvalues of the symmetric stability matrix $(\partial^2 H/\partial u^2)_{mn}$ are non-negative; in fact, these are the squared eigenfrequencies of small Hamiltonian (undamped) oscillations around the equilibrium (u_j). Together with their limit points, they form the *phonon spectrum* of the configuration. The presence of a gap Δ_{ph} in the phonon spectrum, meaning the spectrum is bounded away from zero, reflects the breaking of continuous translational invariance in the lattice state.

Because $h(u, u')$ does not depend on the index (lattice homogeneity) and it is (diagonally) periodic: $h(u + 1, u' + 1) = h(u, u')$, the action of the group of translations $\{\sigma_{rm} : r, m\text{integers}\}$ on configurations (u_j), defined by

[6] *twist* means that $\frac{\partial u_{n+1}}{\partial p_n} > 0$ and this is equivalent to the convexity condition on W.

$$\sigma_{rm}(u_j) = (u_{j+r} + m) \tag{15}$$

preserves the equilibrium property and the minimum energy property.

Given two configurations (u_j), (v_j), one says that the first one is less than the other $(u_j) < (v_j)$ if $u_j < v_j$ for all j. A configuration is *rotationally ordered*, RO for short, if the set of its translates by all the σ_{rm} is a totally ordered set of configurations. It follows that a RO configuration (u_n) has a well defined *mean spacing* between particles (or torsion angle between pendula)

$$\langle u_{n+1} - u_n \rangle = \lim_{n\to\infty, m\to-\infty} \frac{u_n - u_m}{n - m} = \omega \tag{16}$$

and furthermore satisfies

$$|u_{n+r} - u_n - r\omega| < 1 \qquad \text{for all integers } r, n\ . \tag{17}$$

This quite constraining RO property is particularly important, as it turns out that

$$\text{m.e.c.} \Rightarrow \text{RO}\ . \tag{18}$$

Indeed, since the interaction potential $W(\Delta u)$ is convex, it costs less energy for configurations to differ as little as possible (given the potential $V(u)$) from equispaced configurations with same mean spacing, *i.e.* to be RO.

Macroscopic variables are averages over the lattice state, which can be computed with the appropriate distribution functions. For RO configurations, these are the distribution functions (modulo 1), d.f. (mod 1) for short, of real doubly infinite sequences:

Let us denote by $F_{\mathbf{u}}^{M,N}(x)$ the fraction of indices n, $(M \le n \le N)$, for which the fractional part $\mathrm{Frac}(u_n)$ lies in the interval $[0, x)$. The distribution function (mod. 1) $F_{\mathbf{u}}(x)$ of the configuration $\mathbf{u} = (u_j)$, $(0 \le x \le 1)$, is defined as the limit

$$F_{\mathbf{u}}(x) = \lim_{M\to-\infty, N\to\infty} F_{\mathbf{u}}^{M,N}(x) \tag{19}$$

provided the limit exists. Clearly, $F_{\mathbf{u}}(x)$ is a non-decreasing function, with $F(0) = 0$ and $F(1) = 1$. Not every conceivable sequence possesses it, but RO configurations do. Thus, quantities like the mean spacing or the mean energy per particle $\epsilon = \langle h(u_n, u_{n+1}) \rangle$ of RO configurations can be computed as the Stieltjes integrals

$$\omega = \int_0^1 \Delta u \; dF_{\mathbf{u}} \tag{20}$$

$$\epsilon = \int_0^1 h(u, u') \; dF_{\mathbf{u}} \tag{21}$$

where $F_{\mathbf{u}}$ is the d.f. (mod 1) of the configuration (u_j).[7] The d.f. (mod 1) of a RO configuration may either have plateaux at intervals where no particles locate (modulo 1), and jumps, or be a smooth function in some other instances.

For every mean spacing there is at least one recurrent m.e.c. (called *ground state*) [12]. This property (recurrent) means that given any integer r and any number $\epsilon > 0$, there are integers m and $s > 0$ such that both inequalities $|u_{r+s} - u_r - m| < \epsilon$ and $|u_{r+s+1} - u_{r+1} - m| < \epsilon$ are satisfied. Recurrent configurations of the FK model correspond to recurrent orbits of the associated twist map. The distribution function (mod 1) F_ω of a ground state defines a whole class of equivalent ground states as follows: take an arbitrary real α, and consider $f = F_\omega^{-1}$ (the inverse of F_ω) and lift it to the real line by $f(u+1) = f(u) + 1$; then the configuration $(u_n) = (f(n\omega + \alpha))$ is a ground state equivalent to the original one.

Commensurate ground states (u_n) are (spatially) periodic (of minimal period), in the sense that they satisfy a relation of the form $u_{n+q} = u_n + p$ for all n, where $q > 0$ and p are coprime integers, and so their mean spacing is the rational number $\omega = p/q$. These configurations are generically *pinned*, meaning that one has to put some finite energy E_{PN} on the system to displace the configuration over the path-dependent barriers of energy (Peierls-Nabarro barriers) separating ground states[8]. The value of E_{PN} is the energy difference between the ground state and a RO minimax (unstable equilibrium) configuration (the saddle point configuration) ordered between two contiguous ground states.

A pinned configuration has a finite *coherence length* (or *decay range*) of fluctuations, ξ, meaning that if a field component, say u_{n_0}, is externally displaced, this produces displacements of the other components u_n, which decay exponentially with the distance $|n - n_0|$: $\sim \exp(-|n - n_0|/\xi)$. Also, there is a gap $\Delta_{ph} > 0$ in the phonon spectrum.

Ground states with irrational mean spacing ω, called *incommensurate*, can be viewed as limits of sequences of periodic configurations of mean spacing $p_n/q_n \to \omega$ as $n \to \infty$. The physical properties of these structures depend on the parameter K of the model: for each irrational ω there is a critical value $K_c(\omega)$ of K s.t.

- for $K < K_c$ the d.f. (mod 1) is a continuously differentiable strictly increasing function: the family of ground states form a continuum and correspond to the orbits of an invariant Kolmogorov-Arnold-Moser torus for the twist map of the cylinder. The ground state is sliding (unpinned), so that $E_{PN} = 0$. Consequently, localized induced fluctuations extend through the lattice and the decay range diverges: $\xi \to \infty$. The phonon spectrum is gap-less.

[7] In what follows, we will drop the suffix $\mathbf{u}$ when it is clear which configuration we are talking about, or we may use the suffix ω to emphasize that it is the d.f. (mod 1) of a configuration with mean spacing ω.

[8] We assume a gap between contiguous ground states.

- For $K > K_c$ the d.f. (mod 1) becomes the inverse of a so-called Cantor function and the family of ground states form a Cantor set (called Aubry-Mather set, or cantorus in the twist map context). They are pinned: $E_{PN} > 0$, $\Delta_{ph} > 0$, and ξ has a finite value. The transition at $K_c(\omega)$ corresponds to the breaking of the invariant KAM torus into a cantorus, and the fractalization of such continuous invariant sets is a critical phenomenon [13, 14].

Besides the ground states, there are non-recurrent m.e.c. called discommensurations, DC for short. These configurations connect asymptotically ($n \to \pm\infty$) two contiguous commensurate ground states (typically one is a translate of the other, but they can be different in general) by an energy minimizing path of exponentially localized length (the width of the DC) around a lattice site (the center of the DC). They correspond to heteroclinic orbits of the associated area-preserving twist map. The width of the DC is given by the decay range ξ of the recurrent asymptotic configuration. Note also that the d.f. (mod 1) of a DC coincides with that of the recurrent substrate (the deviation of a DC configuration relative to substrate being exponentially localized).

The DCs are elementary defects, i.e., localized compressions (retarded DCs) or expansions (advanced DCs) superimposed on the recurrent substrate modulation. If ω is close to a rational $\omega_0 = p/q$, the ground state with mean spacing ω can be viewed as an "array of DCs" (advanced if $\omega > \omega_0$, retarded if $\omega < \omega_0$) over the recurrent (periodic) ground state of mean spacing ω_0. The interaction energy between neighboring DCs decays exponentially with the quotient inter-distance/width, $\sim \exp(-1/(\xi|\omega - \omega_0|))$, so that for ω very close to ω_0 and/or high values of K (i.e. ξ small), these elementary localized excitations are almost non-interacting (and pinned).

We call Discommensuration Theory the theoretical perspective which describes the modulated phase as a system of localized DCs (weakly, or not so weakly) interacting. Note that it is a description built upon an emergent property which assumes the role of new elementary component of the many-body system, roughly illustrating the notion of *emergent* property in Complex Systems theory [15, 16][9].

There are two limits in which the equilibrium states of a Frenkel-Kontorova model can be found explicitly.

(i) If $K = 0$ (i.e. no external potential), called the *the integrable limit*, then the equation for equilibria reduces to

$$-W'(u_{n+1} - u_n) + W'(u_n - u_{n-1}) = 0 \qquad \text{for all } n \tag{22}$$

that is, the terms $W'(u_{n+1} - u_n)$ are the same for all n. Now the function W' is strictly increasing since $W'' > 0$ (convexity condition), so it follows that for

[9] Temperature, however, is the archetype among the simplest examples, being much richer and deeper.

all n, the difference $u_{n+1} - u_n = u_n - u_{n-1}$ is a constant independent of n, equal to ω, say. Hence equilibria are configurations of the form

$$u_n = n\omega + \alpha \qquad \forall n \tag{23}$$

where α is some constant. So, for every value ω of the mean spacing, there is a one-parameter family (continuum) of equispaced equilibria, parametrized by the phase α.
(ii) If $W(u) \equiv 0$ (i.e. no interaction between neighbors), called the *anti-integrable limit*, then the equation for equilibria reduces to

$$-\frac{\partial H}{\partial u_n} = -V'(u_n) = 0 \qquad \forall n \tag{24}$$

It follows that any sequence of critical points of the potential V is an equilibrium state. For instance, for the Standard FK model, $V'(u_n) = K/2\pi \sin 2\pi u_n = 0$ so any sequence of half-integers is an (anti-integrable) equilibrium state: there are lots of them!

Sequences (u_n) of minima of V are local minima of the total energy and correspond to metastable states. Minimum energy configurations are now only *weakly* rotationally ordered (meaning that translates by the σ_{rm} may touch: $(\sigma_{r,m}\boldsymbol{u})_j \leq u_j$ (or $\geq u_j$) for all j, instead of strict inequalities). In the case where $V(u)$ has a unique minimum per period, at $u = a$, say, it is easy to see that the ground states of mean spacing ω are of the form

$$u_n = \mathrm{Int}(n\omega + \alpha) + a \qquad \forall n \tag{25}$$

where α is a constant. Obviously, this one-parameter family of ground states does not form a continuum like in the integrable limit.

We will use these two limits for our models of collective ratchet transport in further sections.

For a more extensive (and mathematically oriented) presentation of the theory of equilibrium states of FK models, we refer to [17]; the Aubry-Mather theory of minimum energy states can also be found in [18] and [19].

3 Dissipative Dynamics

Non-equilibrium situations in the framework of the FK model are certainly of interest concerning many theoretical and experimental studies. In particular, the behaviour of Hamiltonian nonlinear discrete fields is of fundamental interest in Theoretical Physics. However, in connection to experimental and prospective technological research, two ingredients are often important to add to the ideal Hamiltonian description:

- Coupling to other (external) variables are often amenable to an analysis in terms of driven and damped dynamics and/or
- Thermal effects (Langevin, or other dynamics).

Even neglecting thermal effects (noise), a full characterization of the lattice dynamics governed by equations of the form

$$m\ddot{u}_n + \Gamma\dot{u}_n = C(t)(\Delta^2 u)_n + K(t)V'(u_n) + \mathcal{F}(t) \tag{26}$$

where $(\Delta^2 u)_n$ denotes the discrete Laplacian $u_{n+1}-2u_n+u_{n-1}$ is a formidable task, with too many interesting aspects, which is well beyond the scope of these lectures. We shall restrict here to a particular regime of parameters, the *overdamped* regime[10] (or *dissipative*) limit $m/\Gamma \to 0$, i.e. we shall drop the inertial term $m\ddot{u}_n$ in (26) and consider systems of the form

$$\dot{u}_n = C(t)(\Delta^2 u)_n + K(t)V'(u_n) + \mathcal{F}(t) \tag{27}$$

for which the dynamics are relatively simple due to the following property of (27): if two initial conditions are ordered, their trajectories will remain so at any later time:

$$(u_i(0)) < (v_i(0)) \Rightarrow (u_i(t)) < (v_i(t)) \quad \text{for all } t > 0\,, \tag{28}$$

i.e., the dynamics of the *overdamped* system preserve the (partial) order on sequences [19, 20, 21, 22]. This order-preserving property originates from the convexity of the interaction potential $W(\Delta u)$.

Since the set of translates $\{\sigma_{r,n}(u_j)\}$ of an RO configuration (u_j) is a totally ordered set, we deduce that the RO property is preserved under the dynamics of (27). Given the key role played by rotational order in the theory of m.e. equilibrium states, one may wonder if it can be of use in the analysis of the dynamics of (27). The answer, as will be seen in the lectures by C. Baesens on monotone dynamics, is positive. The dynamics of RO configurations is easier to characterize (the d.f. (modulo 1) $F_{(\tilde{u}_j)}(x;t)$ is well defined and it carries over all the information on the configuration $(\tilde{u}_j(t))$, if it is recurrent). Even more, a RO trajectory $(\tilde{u}_j(t))$ and its translates $\sigma_{r,m}(\tilde{u}_j(t))$ (associated to the same d.f.(mod 1) $F_{(\tilde{u}_j)}(x;t)$) can be used, like a fisher net, to bound (finite spacing) trajectories $(u_j(t))$ that have same mean spacing:

$$\sigma_{r_1,m_1}(\tilde{u}_j(t)) < (u_j(t)) < \sigma_{r_2,m_2}(\tilde{u}_j(t)). \tag{29}$$

This sandwich construction plays a central role in proving, for instance, uniqueness (for initial conditions with bounded spacing and a same mean spacing) of the asymptotic average velocity (or flow) of the chain:

[10] This limit is called *gradient dynamics* in the lectures by C. Baesens. Indeed the right hand side in (27) is the gradient of the total energy H. In those lectures, overdamped means inertial dynamics (26) with ratio m/Γ small enough but not zero.

$$\mathcal{J} = \lim_{T\to\infty} T^{-1} \int_0^T \langle \dot{u}_n \rangle \, dt \tag{30}$$

Extensive numerical studies have motivated (and guided in some cases) rigorous results and theorems in this field. We now highlight some remarkable features of the dynamics of (27) revealed by numerics [21].

1. *Existence of thresholds for transport.* A kind of ubiquitous peculiarity in many observed transport phenomena is the existence of threshold values of the drive parameter, below which equilibrium is stable and no transport occurs. The characterization of thresholds for both cases of additive and parametric driving can be done. For example, under constant additive force $\mathcal{F}(t) = \mathcal{F}$ (and constant parameters K and C) there is a mean spacing dependent depinning force $\mathcal{F}_{th}(\omega) > 0$ if K is large enough, beyond which the structure slides and the asymptotic state is unique [22]. The scaling of different quantities at the depinning transition is nontrivial for irrational values of ω. The issue of thresholds for parametric driving ($\mathcal{F}(t) \equiv 0$, $K(t)$ or $C(t)$ not constant in (27)) will be addressed in the Sect. 6 for a specific model which is amenable to exact analysis.
2. *Multiple attractor coexistence.* The uniqueness of asymptotic sliding states observed for constant additive force does not extend if the additive force is periodic in time. This is also the case for general parametric driving. The abundance of metastable equilibria has its counterpart in the coexistence of multiple resonant (synchronization) states, all sharing the common average velocity value.
3. The addition of a small inertial term $m\ddot{u}_n$ in (26) does not alter qualitatively the dynamics. A preserved partial order has also been found in this case [23].

4 Exercises

1. Distribution functions (mod 1) of the integrable limit configurations.
   (a) Compute the d.f. (mod 1) $F_{\omega,\alpha}(x)$ of the sequence $(\tilde{u}_n) = (n\omega + \alpha)$ for rational $\omega = p/q$ (p and q coprime integers).
   (b) Deduce that for irrational values of ω, $F_{\omega,\alpha}(x) = x$.
2. Use the results of previous exercise to compute the fraction $\mu(\alpha, \beta, \omega)$ of indices n for which $\mathrm{Int}(n\omega + \alpha) \neq \mathrm{Int}(n\omega + \beta)$, for rational and irrational values of ω.

4.1 Solutions to Exercises

1. (a) The sequence $\mathrm{Frac}(np/q + \alpha)$, $-\infty < n < +\infty$, is periodic of period q and takes values $\epsilon + lq^{-1}$, $l = 0 \ldots, q-1$, where $\epsilon = q^{-1}\mathrm{Frac}(q\alpha) < q^{-1}$ for all α. To compute the limit (19) which defines the d.f. (mod 1),

we may consider a subsequence of $F^{M,N}_{\tilde{\mathbf{u}}}(x)$ for which $N-M+1=qj$ ($j=1,2,\dots$), for fixed α and x. Such a subsequence is clearly constant (independent of j), so one only needs to compute $F^{1,q}_{\tilde{\mathbf{u}}}(x)$, this value giving the limit $F_{p/q,\alpha}(x)$.

If $\epsilon < x < 1$, then the number of indices n ($1 \le n \le q$) such that $\mathrm{Frac}(np/q+\alpha) < x$ is the number of consecutive segments of length q^{-1} inside the interval $[\epsilon, x)$, i.e. $1+\mathrm{Int}(q(x-\epsilon)^-)$, where $\mathrm{Int}(x^-)$ denotes the greater integer less than x (i.e. equals the integral part of x, $\mathrm{Int}(x)$, except at integers). This expression gives also the correct (zero) result for $0 \le x \le \epsilon$. Using the property $1+\mathrm{Int}(x) = -\mathrm{Int}((-x)^-)$, one obtains

$$F_{p/q,\alpha}(x) = -q^{-1}\mathrm{Int}(q(-x+q^{-1}\mathrm{Frac}(q\alpha))) \tag{31}$$

Note that $F_{p/q,\alpha}(x+q^{-1}) = F(x)+q^{-1}$, for $x < 1-q^{-1}$. Also, the change $\alpha \to \alpha+q^{-1}$ leaves $F_{p/q,\alpha}(x)$ invariant.

(b) An irrational number ω is the limit of a sequence of rational numbers $p_l/q_l \to \omega$, with $p_l \to \infty$ and $q_l \to \infty$ as $l \to \infty$ (p_l and q_l coprime for each l).

Thus, using the previous result for each rational approximant p_l/q_l and the fact that $q_l^{-1}\mathrm{Frac}(q_l\alpha) \to 0$ for all α as $l \to \infty$, we obtain

$$F_{\omega,\alpha}(x) = \lim_{l\to\infty} -q_l^{-1}\mathrm{Int}(-q_l x) = x \tag{32}$$

which does not depend on α. In other words, orbits of an irrational rotation are uniformly distributed on the circle. A proof of this well-known result, based on the Weyl criterion, can be found in [24].

2. The fraction $\mu(\alpha,\beta,\omega)$ of indices n for which $\mathrm{Int}(n\omega+\alpha) \ne \mathrm{Int}(n\omega+\beta)$ can be obtained from the distribution functions (mod 1) $F_{\omega,\alpha}(x)$ and $F_{\omega,\beta}(x)$ as follows:
Write $n\omega+\beta = n\omega+\alpha+(\beta-\alpha)$ and split terms between their integral and fractional parts, and then realize that

$$\mathrm{Int}(n\omega+\alpha) = \mathrm{Int}(n\omega+\beta) \tag{33}$$

iff

$$\alpha-\beta \le \mathrm{Frac}(n\omega+\alpha) < 1+\alpha-\beta \tag{34}$$

or, equivalently, iff

$$\beta-\alpha \le \mathrm{Frac}(n\omega+\beta) < 1+\beta-\alpha\,. \tag{35}$$

Thus, the indices n violating (33) are those violating any of the inequalities in (34) or, equivalently, any of the inequalities in (35). There are different cases, depending on $\alpha-\beta$:

(i) For $0 \le \alpha-\beta < 1$ the second inequality in (34) holds for all n, while the first inequality is violated when $\mathrm{Frac}(n\omega+\alpha) < \alpha-\beta$,

so $\mu(\alpha,\beta,\omega) = F_{\omega,\alpha}(\alpha-\beta)$. Alternatively, the first inequality in (35) holds for all n, while the second inequality is violated when $\mathrm{Frac}(n\omega+\beta) \geq 1+\beta-\alpha$, so $\mu(\alpha,\beta,\omega) = 1 - F_{\omega,\beta}(1+\beta-\alpha)$.

(ii) For $\alpha-\beta \geq 1$ both inequalities in (34) are violated for all n, so that $\mu(\alpha,\beta,\omega) = 1$.

Hence we obtain

$$\mu(\alpha,\beta,\omega) = \begin{cases} 1 & \text{if } |\alpha-\beta| \geq 1 \\ F_{\omega,\alpha}(\alpha-\beta) = 1 - F_{\omega,\beta}(1+\beta-\alpha) & \text{if } 0 \leq \alpha-\beta < 1 \\ F_{\omega,\beta}(\beta-\alpha) = 1 - F_{\omega,\alpha}(1+\alpha-\beta) & \text{if } 0 \leq \beta-\alpha < 1 \end{cases} \tag{36}$$

where, for rational $\omega = p/q$, $F_{p/q,\alpha}(x)$ is given by (31), and for irrational ω, $F_{\omega,\alpha}(x) = x$ (32). Thus, in the irrational case, $\mu(\alpha,\beta,\omega)$ depends only on the difference $x = \alpha - \beta$, and one obtains the piecewise linear function

$$\mu(x) = \begin{cases} 1 & \text{if } x \leq -1 \\ -x & \text{if } -1 < x \leq 0 \\ x & \text{if } 0 \leq x < 1 \\ 1 & \text{if } 1 \leq x\,. \end{cases} \tag{37}$$

5 Ratchet Effect

Think of a Brownian particle experiencing a mirror-asymmetric periodic potential $V(u)$, so that the density distribution function $\rho(u)$ (d.d.f. for short) of the particle position u is peaked around the minimum of the potential (see Fig. 3). Then turn ideally the potential off, so that the d.d.f. diffuses symmetrically and, after a while, turn the potential on again: as shown in the figure the distribution centroid (or mean position) $\int u\,\rho(u)\,du$ has been shifted. This directional transport results from the symmetric diffusion of the d.d.f. followed by asymmetric localization. The rectification of thermal fluctuations in asymmetric environments is commonly referred to as *ratchet effect* in recent literature [25].

A slightly different version of the ratchet effect is the following: instead of turning on-off the potential $V(u)$, we can monitor the temperature so that the amplitude of the thermal fluctuations (the strength of the diffusive forces) is turned on and off. At zero temperature, the d.d.f. of the particle position is sharply localized at the minimum of the potential, while at high enough temperature, it widens enough to allow ratchet transport when switching temperature on and off cyclically.

Now think of an equilibrium structure of a FK model with a mirror-asymmetric shape for $V(u)$ and turn this potential off, so that only diffusive forces (interaction potential $W(\Delta u)$) act on the lattice. After relaxation to equispaced equilibrium turn the potential $V(u)$ on again. Does one observe transport in this cyclic process? The following unsophisticated argument provides a positive answer to this question (and an estimation of the flow).

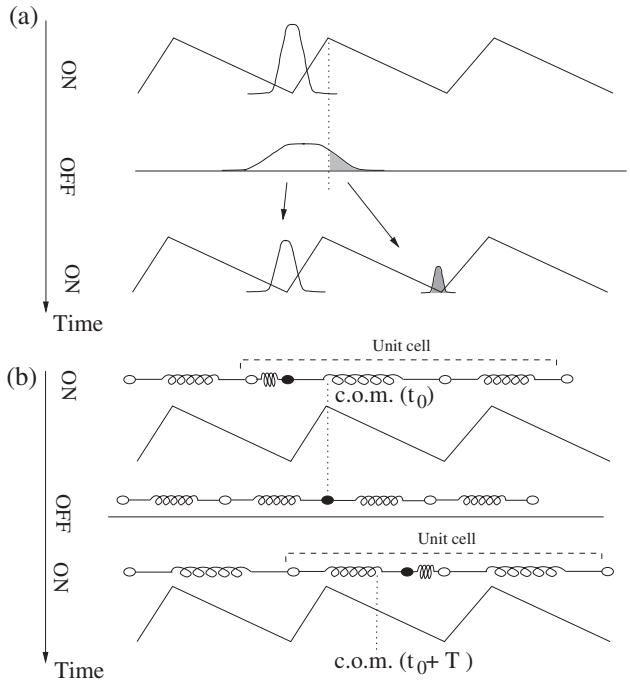


Fig. 3. Thermal versus collective flashing ratchet mechanism. Panel (**a**) shows schematically the thermal broadening of the density distribution function of one particle when the potential is off and how this can produce directed motion when switched on again. In (**b**) the collective interaction between particles in an extended system produces again a net flow when switching on and off the asymmetric potential. In both cases the particle mainly responsible for the flow in this cycle has been highlighted. The center of mass of the unit cell (three particles in two periods of the potential) at t_0 and $t_0 + T$ is marked by the dotted line to see the one-cycle advance

- The distribution centroid $u_{cm}(\mathbf{u}) = \int_0^1 x\, dF_{\mathbf{u}}(x)$, of the configuration $\mathbf{u} = (u_n)$, cannot change when the potential $V(u)$ is turned off, due to Newton's third law. To compute the centroid displacement during the on-semicycle, let us denote by $[\Delta, \Delta + 1]$ the set of values of the initial position $u_0(0)$ of particle 0 in an equispaced configuration, such that $u_0(\infty)$ (the asymptotic position of the particle after the potential has been turned on again) lies inside the 0th well of the potential $V(u)$ (which we assume to be the interval $[0, 1]$). The centroid of the uniform distribution in $[\Delta, \Delta + 1]$ is located at $\Delta + 1/2$, while the distribution centroid of the asymptotic equilibrium is $u_{cm}(bfu(\infty)) = \int_0^1 x\, dF_{\mathbf{u}(\infty)}(x)$. Thus, the centroid displacement in one cycle is simply $u_{cm}(\mathbf{u}(\infty)) - \Delta - 1/2$. One has just to realize that a zero value for this quantity cannot be generic, but indeed rather exceptional provided the mirror-asymmetry condition on $V(u)$.

This "poor man" derivation can be properly formalized, in order to provide a rigorous proof of existence of collective (non thermal) ratchet effect in FK models, as we will see in the following section.

The "temperature on-off" version of the thermal ratchet effect has also a collective ratchet counterpart in the following (non thermal) cycle: Turn on and off the coupling potential $W(u, u')$ of the FK model. For this version we will also find ratchet transport, provided the coupling parameter is turned on above some threshold value which depends on the mean spacing ω (and the specific potential functions).

For both versions of the collective ratchet effect, the Discommensuration Theory [26] provides deep insights into the details of this transport phenomenon. This will be analyzed in Sect. 7, before concluding with a few exercises in Sect. 8.

6 Collective Ratchet Effects in FK Model

To fix ideas, we shall consider here FK models with harmonic coupling, i.e. $W(u, u') = \frac{1}{2}(\Delta u)^2 - \mu \Delta u$, $(\Delta u = u' - u)$, and shall assume that the on-site potential $V(u)$ has a single minimum per period, at $u = a \pmod 1$, $0 < a < 1$, and maxima at integer values of u. More specifically, in the first version of the collective FK ratchet, we will denote by K the amplitude of the on-site potential and write $K\,V(u)$ instead of $V(u)$. For the second version, this amplitude will be kept equal to unity, and we will parametrize the interaction, writing $C\,W(\Delta u)$, instead of $W(\Delta u)$, where C $(= K^{-1})$ is the coupling parameter.

In version 1 (respectively 2), $K(t)$ (resp. $C(t)$) will be a crenelated function of time, taking alternatively values 0 and K (resp. 0 and C), for times of length τ_{off} and τ_{on} respectively, so that the period of a cycle is $T = \tau_{\text{off}} + \tau_{\text{on}}$. Both semi-periods τ_{off} and τ_{on} are assumed to be much larger than the characteristic times of relaxation to equilibrium. This last assumption allows to impose equilibrium conditions on configurations at switching times. We also assume the overdamped limit of dynamics:

$$\dot{u}_n = (\Delta^2 u)_n + K(t)\, V'(u_n) \quad \text{(version 1)} \tag{38}$$

$$\dot{u}_n = C(t)\, (\Delta^2 u)_n + V'(u_n) \quad \text{(version 2)} \tag{39}$$

Under the overdamped dynamics, rotational order and mean spacing are preserved, and we have uniqueness of the asymptotic average velocity (or flow, (30)) for fixed mean spacing; also ground states are RO. These facts will allow us to restrict our analysis to RO configurations for which distribution functions (mod 1) are defined. The algorithm to compute the flow, or equivalently, the displacement of the distribution centroid during a cycle, will be to count the fraction of particles that pass over the position $u = 0 \pmod 1$ during each semicycle, $\mathcal{J}^0_{\text{off}}$ and $\mathcal{J}^0_{\text{on}}$ and add them up.

6.1 Switching Potential $V(u)$ On-off

The initial configuration (before turning the potential off) at $t = 0$ is assumed to be a ground state (i.e., a recurrent and minimum energy configuration) of mean spacing ω for the value K of the potential amplitude. Its d.f. (mod 1) will be denoted by $F_{K,\omega}$ and, using the notation $f = F_{K,\omega}^{-1}$ for the lift to the real line of the inverse of $F_{K,\omega}(x)$, the initial configuration can be written as

$$u_n^\alpha(0) = f(n\omega + \alpha) \tag{40}$$

where α is some constant. Once the potential is turned off, this configuration evolves asymptotically to some equispaced (integrable limit) configuration which we shall assume is reached at $t = \tau_{\rm off}$,

$$u_n^\alpha(\tau_{\rm off}) = n\omega + \beta\ . \tag{41}$$

Given α, the value of β is uniquely determined, due to Newton's third law which imposes that (on average) no net macroscopic motion occurs:

$$\langle u_n^\alpha(\tau_{\rm off}) - u_n^\alpha(0)\rangle = 0\ . \tag{42}$$

By decomposing $u_n^\alpha(0)$ as the sum of its integral and fractional parts, and using $\mathrm{Int}(u_n^\alpha(0)) = \mathrm{Int}(n\omega + \alpha)$ (recall that $F_{K,\omega}(0) = 0$ and $F_{K,\omega}(1) = 1$), (42) yields, for rational $\omega = p/q$ (p, q coprime),

$$\begin{aligned}\beta - \alpha &= q^{-1}\sum_{n=0}^{q-1}(\mathrm{Frac}(u_n^\alpha(0)) - \mathrm{Frac}(n\omega + \alpha)) \\ &= u_{cm}(K, \omega = p/q) - q^{-1}\mathrm{Frac}(q\alpha) - \frac{1}{2}(1 - q^{-1})\end{aligned} \tag{43}$$

where $u_{cm}(K,\omega)$ denotes the d.f. (mod 1) centroid of the configuration $(u_n^\alpha(0))$. Note that the last equality follows from the expression for the distribution centroid in the period-q case, $u_{cm}(K, p/q) = q^{-1}\sum_{n=0}^{q-1}\mathrm{Frac}(u_n^\alpha(0))$, and from the solution to exercise 1 in Sect. 4. For irrational values of ω the result is

$$\beta - \alpha = u_{cm}(K,\omega) - \frac{1}{2}\ . \tag{44}$$

From $0 \le u_{cm} < 1$, one easily deduces that $|\beta - \alpha| \le 1/2 + (2q)^{-1}$, and as[11] $\beta = \alpha$ whenever $q = 1$, we conclude that $|\beta - \alpha| \le 3/4 < 1$ for all q. The consequence of this inequality is that not all particles pass over $u = 0$ (mod 1).

The absolute value of the partial local flow $|\mathcal{J}_{\rm off}^0|$ through $u = 0$ (mod 1) is the fraction $\mu(\alpha,\beta,\omega)$ of indices n such that $\mathrm{Int}(n\omega + \alpha) \neq \mathrm{Int}(n\omega + \beta)$. This quantity has been computed in exercise 2 of Sect. 4 (36–37). Then the flow

[11] Note that the ground states for integer values of ω are already equispaced.

during the off-semicycle is $\mathcal{J}^0_{\text{off}} = \text{sign}(\beta - \alpha)\mu(\alpha, \beta, \omega)$, which we can write as:

$$\mathcal{J}^0_{\text{off}} = \begin{cases} q^{-1}\text{Int}\left(q\left(u_{cm}(K, p/q) - \frac{1}{2}(1 - q^{-1})\right)\right) & \text{for } \omega = p/q \\ u_{cm}(K, \omega) - \frac{1}{2} & \text{for irrational } \omega \end{cases} \tag{45}$$

Consider now the on-semicycle. At the beginning of this semicycle the initial configuration is equispaced, $u^\beta_n(0) = u^\alpha_n(\tau_{\text{off}}) = n\omega + \beta$. Then this initial configuration evolves asymptotically to some recurrent RO equilibrium configuration which we assume is reached at $t = \tau_{\text{on}}$, $u^\beta_n(\tau_{\text{on}})$, a ground state for the parameter values K and ω. In order to compute the partial local flow $\mathcal{J}^0_{\text{on}}$ through $u = 0 \pmod 1$ during this semicycle, we have to introduce a quantity $\Delta(K, \omega)$, which we call the saddle phase. This quantity is the same Δ which appeared in the qualitative argument above in Sect. 5.

Let us denote by M_m the set of values of δ for which $\text{Int}(u^\delta_0(\tau_{\text{on}})) = m$, with m integer, where $u^\delta_n(\tau_{\text{on}})$ denotes the RO configuration of equispaced initial condition $u^\delta_n(0) = n\omega + \delta$. Using the property of order preservation of the dynamics, it is easily seen that if δ_1 and δ_2 are in M_m, then for all λ in $[0, 1]$, $\lambda\delta_1 + (1 - \lambda)\delta_2$ is also in M_m. In other words, M_m is an interval. Moreover, the length of this interval is 1, because if δ is in M_m, then $\delta + 1$ is in M_{m+1}. We define the saddle phase Δ as the infimum of the interval M_0.

If it is the case that the ground state equilibria corresponding to (K, ω) are pinned configurations, they are attracting fixed points of the dynamics, each one surrounded by its basin of attraction, the open set of initial configurations asymptotically evolving to the fixed point. The frontier of two contiguous basins contains an unstable equilibrium, a saddle (or minimax) configuration, which attracts only initial conditions on the frontier. On the other hand, the equispaced equilibria $(n\omega + \delta)$ of the system at $K = 0$ form a line in the phase space, parametrized by δ. The intersection of the stable manifold of the saddle (*i.e.* the frontier) with the line of equispaced configurations is just $(u^\Delta_n(0))$. This geometrical interpretation justifies the term "saddle phase" for Δ. In a more physical language, the Peierls barrier (which is the energy of the saddle configuration relative to stable equilibria), produces the opening of a step (often called Peierls gap) in the distribution function (mod 1), which starts at Δ. If the equilibria are not pinned, then Δ corresponds to the phase of the initial configuration such that $u^\Delta_0(\tau_{\text{on}})$ lies on top of a potential maximum. Though one cannot speak of a saddle in this case, we will keep the term for Δ.

Before computing the partial local flow $\mathcal{J}^0_{\text{on}}$, we now prove that $|\Delta| < 1$ (The following notation is used, $u^{\delta^+}_n = \lim_{\delta \to \delta^+} u^\delta_n$, and $u^{\delta^-}_n = \lim_{\delta \to \delta^-} u^\delta_n$). If it were the case that $\Delta \leq -1$, then one could always find an RO equilibrium configuration $(\tilde{u}_n)$ with $\text{Int}(\tilde{u}_0) = -1$ and $u^{\Delta^+}_n < \tilde{u}_n$ for all n (because the fluctuations of a RO configuration with respect to equispacing are strongly bounded – recall (17)). On one hand, as $(\tilde{u}_n)$ is an equilibrium configuration, $\tilde{u}_n(\tau_{\text{on}}) = \tilde{u}_n(0)$ for all n, so that $\text{Int}(\tilde{u}_0(\tau_{\text{on}})) = -1$; on the other hand, order preservation implies that $u^{\Delta^+}_n(\tau_{\text{on}}) \leq \tilde{u}_n(\tau_{\text{on}})$, so that $0 = \text{Int}(u^{\Delta^+}_n(\tau_{\text{on}})) \leq$

$\mathrm{Int}(\tilde{u}_n(\tau_{\mathrm{on}})) = -1$. This contradiction proves that $\Delta > -1$. *Mutatis mutandis*, the same type of argument also proves that $\Delta < 1$.

As the initial configuration is $u_n^\beta(0) = n\omega + \beta$, one easily realizes that the fraction of particles passing over $u = 0 \pmod 1$ during the on-semicycle is the fraction of indices n such that

$$\begin{aligned} \mathrm{Frac}(n\omega + \beta) < \Delta \qquad & \text{if } \Delta > 0 \\ \mathrm{Frac}(n\omega + \beta) \geq 1 + \Delta \qquad & \text{if } \Delta \leq 0 \end{aligned} \tag{46}$$

that is, $\tilde{F}_{\omega,\beta}(\Delta)$ if $\Delta > 0$, and $1 - \tilde{F}_{\omega,\beta}(1+\Delta)$ if $\Delta > 0$, where the d.f. (mod 1) $\tilde{F}_{\omega,\beta}(x)$ was computed in Exercise 1 of Sect. 4 (expressions (31) and (32)). Note that particles cross $u = 0 \pmod 1$ from right to left if $\Delta > 0$ (negative flux) and from left to right if $\Delta < 0$ (positive flux). Then using the expressions (43) and (44) for $\beta - \alpha$, we obtain

$$\mathcal{J}^0_{\mathrm{on}} = \begin{cases} q^{-1}\mathrm{Int}\big(- q\Delta + \mathrm{Frac}\big(q\big(u_{cm} - \frac{1}{2}(1 - q^{-1})\big)\big)\big) & \text{for } \omega = p/q \\ -\Delta & \text{for irrational } \omega \end{cases} \tag{47}$$

(the same expressions for both signs of Δ).

Finally, the flow $\mathcal{J}(K,\omega)$ during the cycle is the sum of (45) and (47):

$$\mathcal{J}(K,\omega) = \begin{cases} \frac{1}{q}\ \mathrm{Int}\big(q\big(u_{\mathrm{cm}}\big(K, \frac{p}{q}\big) - \Delta\big(K, \frac{p}{q}\big) - \frac{1}{2}\big(1 - \frac{1}{q}\big)\big)\big) & \text{for } \omega = p/q \\ u_{\mathrm{cm}}(K,\omega) - \Delta(K,\omega) - \frac{1}{2} & \text{for irrational } \omega\ . \end{cases} \tag{48}$$

Both quantities, the distribution centroid $u_{cm}(K,\omega)$ and the saddle phase $\Delta(K,\omega)$, are continuous functions of K and ω. Indeed u_{cm} is a ground state average, and the Aubry-Mather theory assures that the ground state d.f. (mod 1) changes continuously. On the other hand, the saddle configuration also changes continuously, so that one expects the saddle stable manifold and consequently the saddle phase to vary continuously as well.

Thus, for the version 1 of the collective ratchet effect in the standard FK model, the flow is a well defined continuous function $\mathcal{J}(K,\omega)$ with point discontinuities at rational values of $\omega = p/q$; the size of these discontinuities behaves as $(2q)^{-1}$.

Given a particular potential function $V(u)$, the quantities $u_{cm}(K,\omega)$ and $\Delta(K,\omega)$ determining the flow can be computed with arbitrary numerical precision: The computation of u_{cm} only requires the obtention of the corresponding ground state configuration, which can be efficiently achieved by numerical integration of (38) for an equispaced initial configuration. Combining this with a simple bisection method provides also the value of Δ. Numerical results [27] of these quantities for an arbitrary chosen $V(u)$ are shown in Figs. 4(1) and 4(2). The resulting flow $\mathcal{J}(K,\omega)$ is shown in Fig. 4(3).

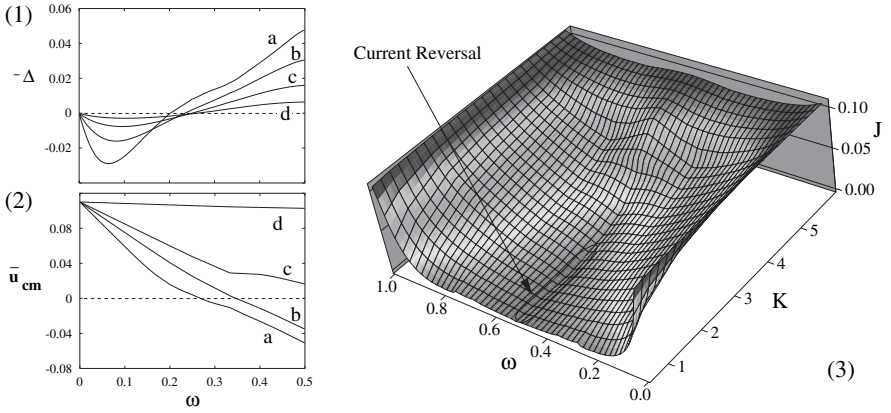


Fig. 4. Numerical results using the pinning potential $V(u) = (2\pi)^2[\sin(2\pi(u+b)) + 0.22\sin(4\pi(u+b))]$ with $b = 0.194969$. (**1**) shows the saddle phase $\Delta(K,\omega)$ for $K = 1, 2, 4, 10$ (labeled from a to d). The mean asymmetry $\overline{u}_{cm} = u_{cm} - 1/2$ is plotted in (**2**) for the same values of K. (**3**) shows the surface $\mathcal{J}(K,\omega)$ for $0.5 < K < 6.0$ and $0 < \omega < 1$. The figure reflects the crests of high flow near integer commensuration and the small region of current reversal around $\omega = 0.5$ at low values of K

6.2 Switching Interaction $W(\Delta u)$ On-off

We now consider the version 2 of the collective ratchet effect in the FK model. The initial configuration (before turning the coupling off) is a ground state $u_n^\alpha(0) = f(n\omega + \alpha)$. When the interaction $W(\Delta u)$ is switched off (anti-integrable limit), every single particle in the configuration evolves towards the minimum of its potential well, so that the asymptotic configuration is a recurrent stable RO uncoupled configuration, which can be written as

$$u_n^\alpha(\tau_{\rm off}) = {\rm Int}(n\omega + \alpha) + a. \tag{49}$$

Thus, during the off-semicycle, no particles pass over $u = 0$ (mod 1), and the partial local flow $\mathcal{J}^0_{\rm off} = 0$.

Now the coupling parameter is turned on to the value C. It is easy to realize that for small enough values of C no particles can pass over a potential maximum. Indeed, if the coupling parameter is much smaller than the (absolute value of the) maximal slope of the pinning potential $V(u)$, the coupling forces cannot overcome the pinning forces and each particle will remain in its initial potential well, so that $\mathcal{J}^0_{\rm on} = 0$ for all ω. This suggests the existence of a threshold value of the coupling, though it may depend on ω.

In order to prove that there is a threshold value, likely dependent on the mean spacing ω, $C_{\rm th}(\omega) < \infty$, such that for $C > C_{\rm th}(\omega)$ the flow is non-zero, we can analyze the limit $C \to \infty$ of version 2. This limit is equivalent to the limit $K \to \infty$ of the version 1; in this limit the distribution centroid u_{cm} of

uncoupled configurations tends to a for all ω, and the saddle phase Δ tends to zero, so that from the (48) one obtains the limit flow

$$\mathcal{J}(\infty,\omega) = \begin{cases} q^{-1}\mathrm{Int}\left(q\left(a-\frac{1}{2}(1-q^{-1})\right)\right) & \text{if } \omega = p/q \\ a-\frac{1}{2} & \text{if } \omega \text{ is irrational .} \end{cases} \tag{50}$$

One concludes that, provided $a \neq 1/2$, the flow is zero only for those rationals $\omega = p/q$ such that $q \leq (2|a-1/2|)^{-1}$, which is a finite set of values in the unit interval. Then, for each mean spacing ω (except at most a finite number of rationals per unit interval), there is a threshold value $0 < C_{\mathrm{th}}(\omega) < \infty$ of the coupling parameter separating two different regimes.

In order to visualize the existence of thresholds for transport, one can think of the phase portrait in configuration space: The recurrent stable RO equilibrium $u_n = f(n\omega+\alpha)$ has a basin of attraction whose frontier is the saddle stable manifold. For low values of C, the uncoupled configuration $\mathrm{Int}(n\omega+\alpha)+a$ belongs to this basin of attraction. Increasing the value of C results in a continuous variation of the ground state, the saddle, and the saddle stable manifold. When $C = C_{\mathrm{th}}(\omega)$ the saddle stable manifold touches on the point in configuration space representing the uncoupled configuration. A further increase of C places the uncoupled configuration inside the basin of attraction of the contiguous ground state.

We show in Fig. 5 the numerical computations [28] of $C_{\mathrm{th}}(\omega)$ for the same potential function $V(u)$ as in Fig. 4. One can clearly observe there both jump and point discontinuities at rational values of ω. Note that $a \approx 0.610061$ and $(2|a-1/2|)^{-1} \approx 4.5$ for this potential, so C_{th}^{-1} could only be zero for $q \leq 4$, as confirmed by the numerics. Figure 6 shows numerical computations of the flow $\mathcal{J}(C,\omega)$ for some values of the coupling parameter C; the most remarkable aspect of these graphs is their piecewise linear shape. In the next section we will discuss these features from the perspective of the Discommensuration Theory.

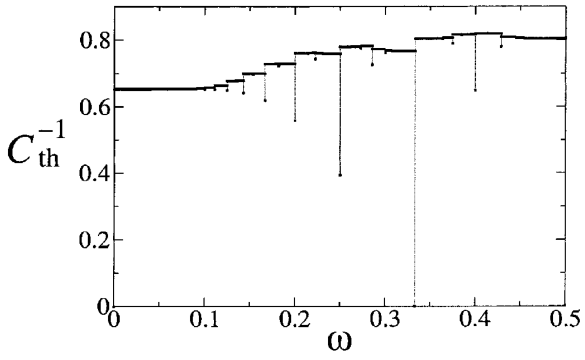


Fig. 5. Numerical computation of the inverse of the threshold coupling parameter (C_{th}^{-1}) as a function of ω (*drop lines* are guides for the eyes)

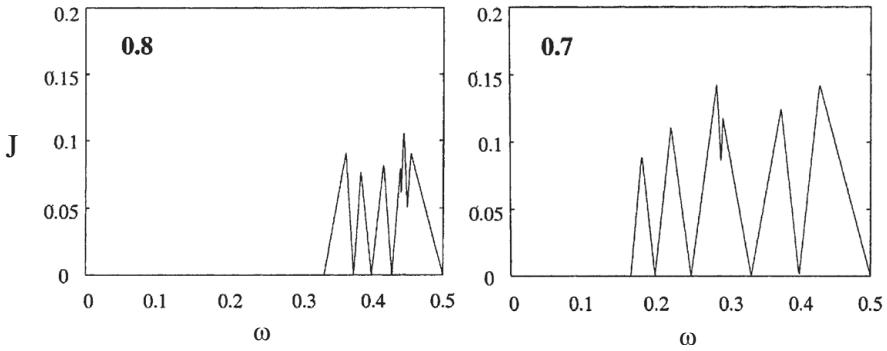


Fig. 6. Plot of the flow $\mathcal{J}(\omega)$ for two different values of the inverse coupling parameter $C^{-1} = 0.8$ and 0.7

7 Discommensuration Theory

As mentioned in Sect. 2, according to the Discommensuration (DC) Theory, a ground state of mean spacing ω, with phonon gap, is well approximated by an "array of DCs" on a ground state of close enough rational mean spacing $\omega_0 = p/q$, more precisely, by a concatenation of segments of DCs of mean length $|q\omega - p|^{-1}$, with advanced DCs if $\omega > \omega_0$ and retarded DCs if $\omega < \omega_0$, the error being exponentially small in $|q\omega - p|$: at most $\exp(-1/(2\xi|q\omega - p|))$, where ξ is the decay range (coherence length) of the DC.

So provided ξc is small, where $c = |q\omega - p|$ is the density of DCs (inverse of the average spacing between DCs), the DCs in the array are almost non-interacting (thus almost identical) field excitations. This allows to describe the physical properties of the ground state of mean spacing ω in terms of the DC properties relative to the substrate configuration (the ground state of mean spacing ω_0).

A convenient variable for the representation of a DC is the averaged relative positions φ_j^{DC} of the DC configuration (w_j^{DC}) with respect to the substrate configuration $(u_j^{\omega_0})$:

$$\varphi_j^{\mathrm{DC}} = q^{-1} \sum_{i=0}^{q-1} \left(w_{j+i}^{\mathrm{DC}} - u_{j+i}^{\omega_0} \right) \ . \tag{51}$$

Figure 7 illustrates the "intrinsic" properties of a DC: The *excess length* of a DC is the asymptotic difference $\lim_{j\to\infty} \varphi_j^{\mathrm{DC}} - \lim_{j\to-\infty} \varphi_j^{\mathrm{DC}} = \pm q^{-1}$, where $+$ is for advanced and $-$ for retarded DCs. The *width* of a DC is twice the decay range ξ of the substrate configuration, and the *center* of a DC is an index j_0 for which the deviation from the substrate configuration is maximum. A

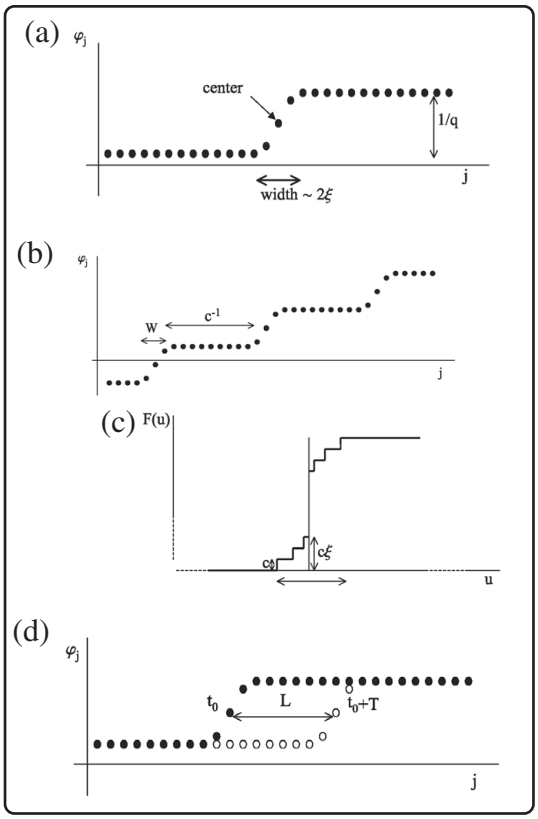


Fig. 7. (**a**) Graphical representation of the intrinsic properties of an advanced DC (see text for definitions). (**b**) Array of equispaced DCs. (**c**) Distribution function (mod 1) of an array of equispaced DCs over a commensurate configuration: each jump in the d.f. (mod 1) of the commensurate configuration (central vertical line in the figure) is now splitted into several jumps at the DC particle positions, of height c. (**d**) The relative advance L of the DC center during a time T involves a microscopic total displacement of L/q relative to substrate

finite displacement of the DC center $j_0 \to j_0 + L$ involves a total displacement of particles $\approx \mp Lq^{-1}$ relative to the substrate.

An array of identical DCs (on a substrate of mean spacing ω_0) with density (inverse number of particles between DC centers) $c = q|\omega - \omega_0| = |q\omega - p|$ has a d.f. (mod 1) F_ω as sketched in Fig. 7.c: there are new jumps (with respect to the substrate d.f. F_{ω_0}) located at the DC particle positions w_j^{DC} (mod 1), and the height of these jumps is the density c of the array. From these observations one can see that the F_ω-average of a period-1 test function g can be expressed as

$$\int_0^1 g\, dF_\omega = \int_0^1 g\, dF_{\omega_0} + c\, g^{\mathrm{DC}} \tag{52}$$

where the quantity g^{DC} is a characteristic (intrinsic property) of a single DC, thus independent of ω. The importance of (52) is that it provides the full Taylor (power) series expansion of macroscopic variables in a neighborhood of ω_0 (note that $c = q|\omega - \omega_0|$), which contains no terms higher than linear.

We will now derive explicitly (52) for the particular case $g(u) = \mathrm{Frac}(u)$, i.e. compute $u_{\mathrm{cm}}(\omega) - u_{\mathrm{cm}}(\omega_0)$ for values of ω very close to the rational ω_0:

$$u_{\mathrm{cm}}(\omega) - u_{\mathrm{cm}}(\omega_0) = \lim_{N\to\infty} (2N+1)^{-1} \sum_{n=-N}^{N} (\mathrm{Frac}(u_n^{\omega}) - \mathrm{Frac}(u_n^{\omega_0})) \quad (53)$$

$$\approx \lim_{N\to\infty} (2N+1)^{-1} \sum_{n=-N}^{N} (\mathrm{Frac}(v_n) - \mathrm{Frac}(u_n^{\omega_0})) \quad (54)$$

where (v_n) is the configuration of concatenated DCs, and an error of size at most $\exp(-1/(2\xi|q\omega - p|))$ is incurred. Now, assuming $2N+1 \gg c^{-1} \gg 1$ we approximate the sum above as the number $c(2N+1)$ of DCs in the segment of length $2N+1$ times the average contribution of each DC. To make sense of the contribution $\sum_{n=-M}^{N}(\mathrm{Frac}(w_n^{\mathrm{DC}}) - \mathrm{Frac}(u_n^{\omega_0}))$ of a DC, where w_n^{DC} denote the particle positions in a DC configuration centered at $n = 0$, one has to consider the limit of the sum over whole numbers of period q, $\lim_{k\to-\infty,l\to+\infty} \sum_{n=kq+i}^{lq+i}(\mathrm{Frac}(w_n^{\mathrm{DC}}) - \mathrm{Frac}(u_n^{\omega_0}))$, else in general it oscillates for ever as $M, N \to \infty$. This still depends on the phase $i \pmod q$, but it can be shown to cycle periodically through the q possibilities along the array of DCs. So the end result is that

$$\lim_{N\to\infty} (2N+1)^{-1} \sum_{n=-N}^{N} (\mathrm{Frac}(v_n) - \mathrm{Frac}(u_n^{\omega_0})) \approx c\,\alpha^{DC}\,, \quad (55)$$

with

$$\alpha^{DC} = \frac{1}{q} \sum_{i=1}^{q} \left(\sum_{k=-\infty}^{\infty} \left(\sum_{n=kq+i}^{kq+i+q-1} (w_n^{\mathrm{DC}} - u_n^{\omega_0}) \right) \right). \quad (56)$$

We can thus write, with exponentially small error:

$$u_{\mathrm{cm}}(\omega) = u_{\mathrm{cm}}(\omega_0) + |\omega - \omega_0| q \alpha^{\mathrm{DC}} \quad (57)$$

where α^{DC} can be interpreted as the relative asymmetry of the DC with respect to the substrate ($\omega_0 = p/q$) configuration. In formula (57), it is to be understood that α^{DC} must be computed with the advanced DC if $\omega > p/q$ and with the retarded one if $\omega < p/q$.

From (57), we see that whenever $\alpha^+ \neq -\alpha^-$ ($\pm$ denoting advanced/retarded DC), there is a jump discontinuity in the first derivative of $u_{\mathrm{cm}}(\omega)$. Unless some special symmetry is invoked in particular cases, these singularities must occur generically. Indeed, they are easily observed in the numerical computations [27] of the distribution centroid shown in Fig. 4.b. We also see

clearly in Fig. 4.b that the saddle phase $\Delta(\omega)$ has a discontinuous first derivative. However this quantity is not a F_ω-average, so that the explanation of this observation cannot be made on the basis of (52) (note also that the very concept of DC looses all meaning in the integrable limit $K = 0$).

As one should not expect that the discontinuities in the slope of $u_{\rm cm}(\omega)$ exactly cancel those of $\Delta(\omega)$, one concludes that the flow $\mathcal{J}(K, \omega)$ shows discontinuous first partial derivative with respect to ω at rationals. Indeed, our numerical results are fully consistent with this expectation.

For the version 2 of the collective ratchet effect, the DC theory remains a valid approach during the complete cycle, provided C is bounded. An interesting consequence for the functional form of the flow dependence on ω is the following:

Assume that ω is close to a rational $\omega_0 = p/q$, $\omega > \omega_0$ (respectively $\omega < \omega_0$), and $C_{\rm th}(\omega) < C < \infty$. Let r^+ (resp. r^-) be the displacement of the advanced (resp. retarded) DC center relative to the substrate after a complete on-off cycle. (Note that, in general, r^+ and r^- can be different integers.) The flow $\mathcal{J}(C, \omega)$ can thus be written as

$$\begin{aligned} \mathcal{J}(C, \omega) &= \mathcal{J}(C, \omega_0) \mp r^{\pm} |\omega - \omega_0| \\ &= \mathcal{J}(C, \omega_0) - r^{\pm}(\omega - \omega_0) \end{aligned} \tag{58}$$

with exponentially small error. On one hand, this assures continuity of the flow respect to ω, and on the other a discontinuous first partial derivative whenever $r^+ \neq r^-$. The numerical computations of the flow shown in Fig. 6 confirm these expectations.

From the DC theory basic tenet, a DC configuration is the one-sided limit of a sequence of (commensurate or incommensurate) ground state configurations of average spacings ω_j approaching the rational p/q from the left (retarded DC) or right (advanced DC) side. Thus, the one-sided limit $C_{\rm th}((p/q)^{\pm})$ is the threshold coupling for the (advanced/retarded) DC configuration at $\omega = p/q$. Unlike the centroid, the threshold coupling is not a configuration average, and then there is no reason that $C_{\rm th}(p/q)$, $C_{\rm th}((p/q)^+)$ and $C_{\rm th}((p/q)^-)$ should coincide and they indeed do not, in general, as shown by numerical results in Fig. 5.

8 Exercises

1. A close inspection of the graph of $C_{\rm th}^{-1}(\omega)$ in Fig. 5 reveals that the value of $C_{\rm th}(\omega)$ at rationals is greater than its left-hand and right-hand limits there and, apart from these point discontinuities, $C_{\rm th}(\omega)$ is a step function. Use Discommensuration Theory to explain these numerical observations.
2. Metastable configurations of the FK model are stable configurations which are not m.e.c. They are assured to exist for all values of ω, provided K is high enough (or, equivalently, C is low enough). A simple type of

metastable structure is an array of DCs on a commensurate substrate with a non-homogeneous inter-spacing between contiguous DCs. For the version 2 of the collective ratchet effect, an initial metastable configuration of this type can evolve after a complete on-off cycle into a translate of itself.
   (a) Give an argument supporting the previous statement.
   (b) Work out a numerical example confirming this assertion.
   (c) Argue that this is not possible to happen for the version 1 of the collective ratchet effect.

3. Along these lecture notes we have assumed thermodynamic limit conditions. An important issue in some applications to experimental systems is how finite size effects can modify the system behaviour. Concerning dynamics, in particular, it is important to note that for finite chains the mean spacing ω is no longer a constant of motion. The parameter which controls the finite system length in the Standard model ($W(\Delta u) = \frac{1}{2}(\Delta u)^2 - \sigma \Delta u$) is the tension σ (i.e. the unstretched length of the "springs"). Also, the very notion of RO configuration is not so useful for a finite system. The aim of this exercise is to illustrate some of the differences in the behaviour in the version 1 of the ratchet collective effect. Consider an equispaced initial configuration of 720 particles with spacing 3/80, and integrate the equations of motion of overdamped dynamics at $K = 0.05$ for both periodic boundary conditions (pbc – infinite system size) and a finite system with $\sigma = 3/80$ with free boundary conditions (fbc).
   (a) Compute the flow in both cases and plot the φ-profile:

$$\varphi_j = q^{-1} \sum_{i=0}^{q-1} \left(u_{j+i}^{\mathrm{fbc}} - u_{j+i}^{\mathrm{pbc}} \right) . \tag{59}$$

   (b) Explain the φ-profile and the difference in the flow in terms of the asymmetric entrance of DCs at boundaries.
   (c) Answer question (a) for $K = 0.4$, and comment on the differences with respect to the case $K = 0.05$.

8.1 Solutions to Exercises

1. The positions (mod 1), $\mathrm{Frac}(w_j^{\mathrm{DC}})$, of the particles around the DC center j_0 are placed inside the gaps between the particle positions (mod 1) of the ground state configuration (see Fig. 7); in particular, there are particles in the DC configuration closer to the potential maxima. Then one expects that the threshold coupling of a DC is lower than the threshold of the commensurate substrate: $C_{\mathrm{th}}(p/q) > C_{\mathrm{th}}((p/q)^{\pm})$.

   For the second observation, to fix ideas, consider an array of advanced DCs ($\omega >\sim p/q$) with a very low density $c = q\omega - p$, so that the DCs behave independently, i.e. each of them behaves as a single DC does. For $C < C_{\mathrm{th}}((p/q)^{+})$ neither the DCs nor the substrate can move, then $C_{\mathrm{th}}(\omega) \geq C_{\mathrm{th}}((p/q)^{+})$ for $\omega >\sim p/q$. On the other hand, as the DCs

are assumed independent, they do move for $C > C_{\rm th}((p/q)^+)$, so that $C_{\rm th}(\omega) \leq C_{\rm th}((p/q)^+)$. Thus, provided the assumption of independent DCs holds, one concludes $C_{\rm th}(\omega) = C_{\rm th}((p/q)^+)$ for $\omega >\sim p/q$.

2. (a) Provided the distances between contiguous DCs in the array are large enough, each DC will move independently of the others for values of the coupling larger than the DC threshold. After a complete on-off cycle, the displacement of each DC center will be the same, then the distances between DCs will remain as the initial ones, and the final configuration will be a translate of the initial metastable configuration.

   (b) First, one has to compute a DC configuration on a commensurate ground state. We make the arbitrary choice $\omega_0 = 1/2$ and compute the advanced DC configuration by integrating the overdamped equations of motion (using a standard Runge-Kutta algorithm) for an initial equispaced configuration of 251 particles over 125 periods of $V(u)$. We use the same potential function as in the numerical examples shown in previous sections, and $C = 2$. Then, we construct the initial metastable structure by joining four finite pieces (of different length) of the DC configuration centered around the DC center and finally impose periodic boundary conditions (see Fig. 8).

   After numerical integration of the equation of motion (during semicycle times large enough to assure that equilibrium is effectively reached after each semicycle), we observe that the final configuration is a translate of the initial one (see Fig. 9). As expected from the argument above, each DC center has been shifted by the same amount.

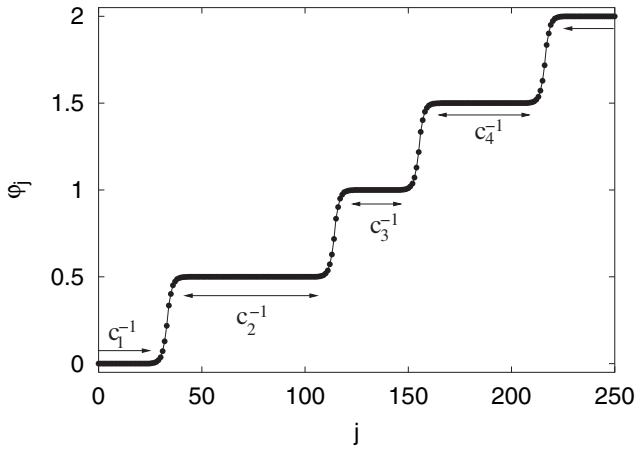


Fig. 8. Plot of φ_j for an array of 4 discommensurations. The substrate is an $\omega_0 = 1/2$ configuration, and $\omega = 21/40$. The coupling parameter is set to $C = 2$. The separation between the DCs (c_i^{-1}) in the metastable configuration are symbolized by the arrows between them

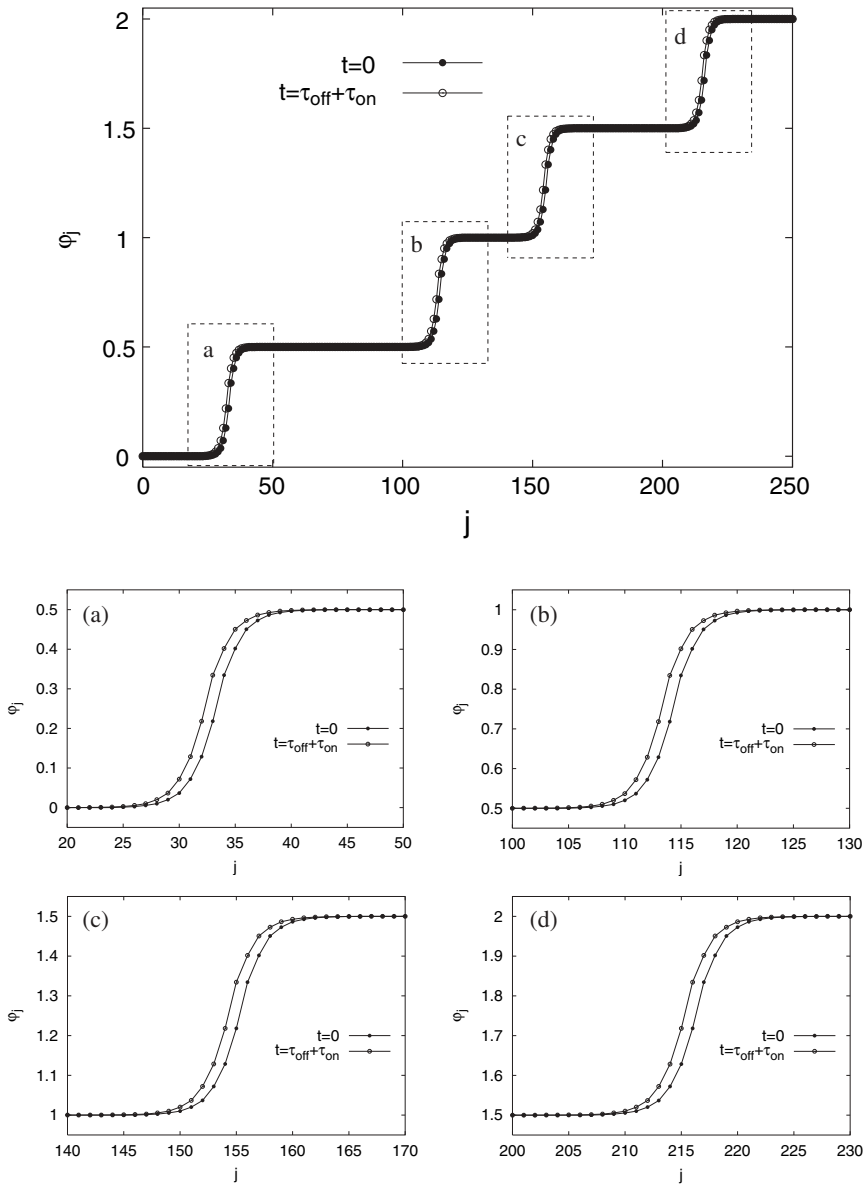


Fig. 9. Evolution of the metastable configuration of Fig. 8 after one period $\tau_{\rm off}+\tau_{\rm on}$. The initial distance between the DCs is preserved. As can be seen from the initial and final state of each DC ((**a**), (**b**), (**c**) and (**d**)) all the DCs centers have been shifted 1 site to the *left*

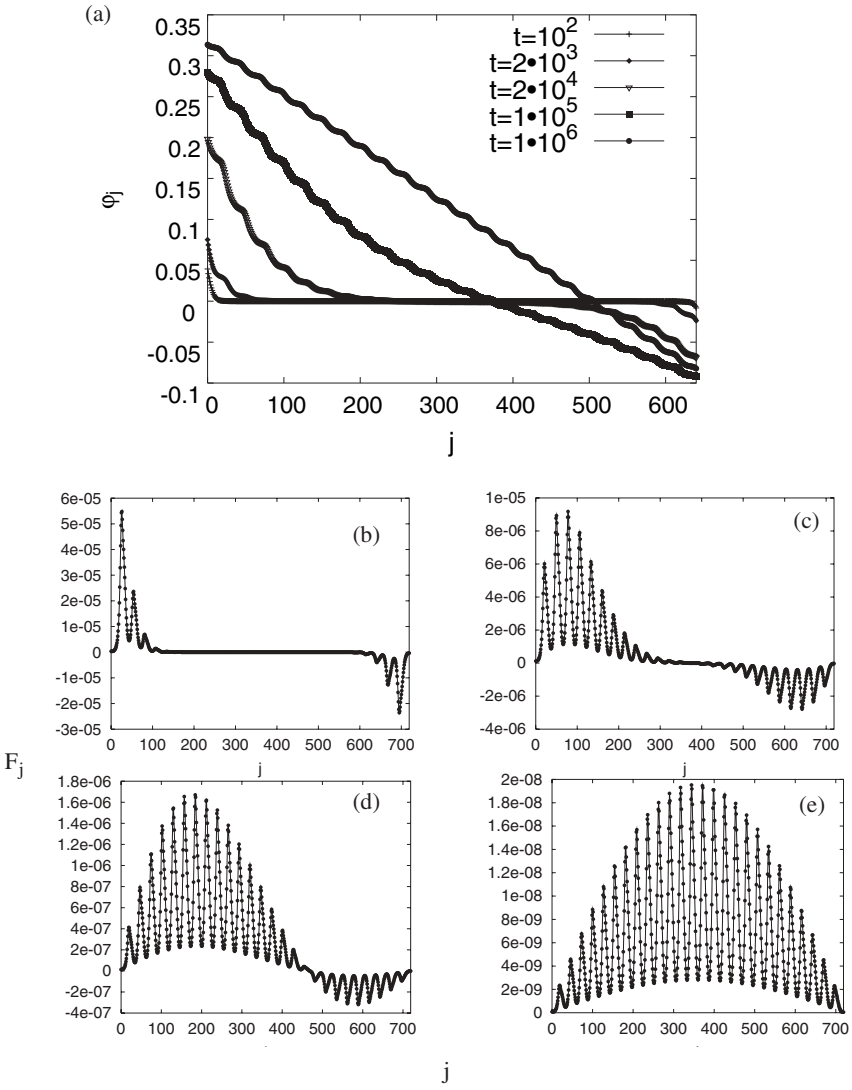


Fig. 10. (**a**) Time evolution of φ_j. The times of the instantaneous profiles are $t = 10^2$, $2 \cdot 10^3$, $2 \cdot 10^4$, 10^5 and $10^6 \simeq \tau_{relax}$. Figures (**b**), (**c**), (**d**) and (**e**) show the instantaneous force (or velocity) distribution of the fbc configuration at $t = 2 \cdot 10^3$, $2 \cdot 10^4$, 10^5 and 10^6 respectively

(c) In the version 1 of the collective ratchet effect, any initial metastable configuration evolves asymptotically during the off-semicycle to an equispaced configuration. This cannot evolve back to metastable during the on-semicycle, so that the dynamical persistence of metastable configurations is impossible in version 1.

3. (a) The results of the numerical computation of the φ-profile at different evolution times are plotted in Fig. 10. The flow difference is $0.12174 - 0.00589 = 0.11585$.

   (b) We observe a negative average slope of the φ-profile. Thus, the asymptotic equilibrium of the fbc configuration has a shorter total length than the pbc equilibrium; it can be described as an array of (equispaced) retarded DCs over the pbc configuration.

   The negative excess length (which for initial times is concentrated at the boundaries) spreads slowly over the whole chain. In other words, DCs enter on a short time scale, and then propagate on a larger time scale toward inner regions. One also observes that more DCs enter from the left boundary than from the right one, so that the center of mass of the final fbc configuration is displaced relative to that of the pbc one, explaining the flow difference. In Figs. (b) to (e) we plot the instantaneous force (or velocity) profile of the fbc chain.

   (c) For $K = 0.4$ the flow difference is much smaller: $0.04911 - 0.04653 = 0.00258$. As observed in the φ- profile in Fig. 11, the DCs that enter from the boundaries (due to the shortening of the chain length) are unable to overcome the Peierls-Nabarro barrier to their motion, and the negative excess length remains concentrated near boundaries. Consequently, inner regions of the chain do not feel any effect from the boundaries, and the macroscopic behaviour is unaffected by boundary conditions.

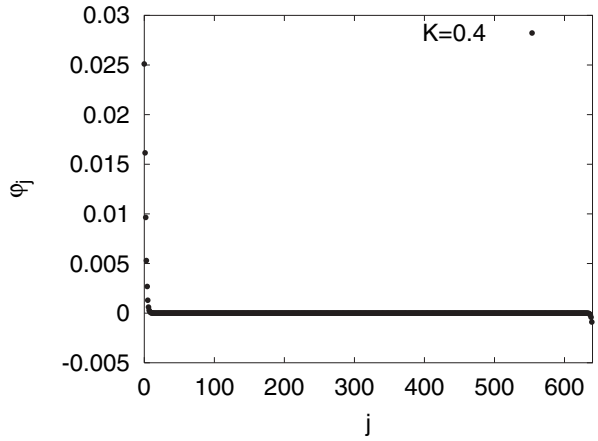


Fig. 11. Profile of φ_j for the $K = 0.4$ free ends configuration with $\sigma = 3/80$ and $N = 720$

References

1. S.J. Gould *The Structure of Evolutionary Theory* (Harvard University Press, Cambridge 2002). 209
2. P.M. Chaikin and T.C. Lubensky *Principles of condensed matter physics* (Cambridge University Press, Cambridge 1995). 210
3. O.M. Braun and Yu. S. Kivshar, *The Frenkel-Kontorova model* (Springer, New York 2004). 210
4. D.K. Campbell, S. Flach, and Yu. S. Kivshar, *Phys. Tod.* **57**, 43 (2004). 211
5. J.J. Mazo, in *Energy Localisation and Transfer*, edited by T. Dauxois, A. Litvak-Hinenzon, R.S. Mackay and A. Spanoudaki (Singapore: World Scientific) 2003. 211, 212
6. S. Watanabe, H.S.J. van der Zant, S.H. Strogatz, and T. Orlando, *Physica D* **97**, 429 (1996). 213
7. F. Falo, P.J. Martínez, J.J. Mazo, T.P. Orlando, K. Segall and E. Trías, *Appl. Phys. A* **75**, 263 (2002). 213
8. M. Peyrard and A.R. Bishop, *Phys. Rev. Let.* **62**, 2755 (1989). 214
9. Th. Dauxois, M. Peyrard and A.R. Bishop *Phys. Rev. E* **47**, 684 (1993). 214
10. J.D. Weeks et al. arXiv:cond-mat/0406246v1 (2004). 214
11. M. Ridley *Evolution (second edition)* (Blackwell, Cambridge 1996). 214
12. S. Aubry and P.Y. Le Daeron, *Physica D* **8**, 381 (1983). 217
13. S. Aubry , in *Structures et Instabilités*, edited by C. Godrèche (Les Ulis, France: Editions de Physique) 1985. 218
14. R.S. Mackay, Ph.D. thesis, Princeton University. Reprinted version in: R.S. Mackay, *Renormalisation in Area-preserving Maps*, (Singapore: World Scientific) 1993; *Physica D* **50**, 71 (1991). 218
15. A. Scott, *Nonlinear Science: Emergence and Dynamics of Coherent Structures*, (Oxford: Oxford University Press) 1999. 218
16. A. Scott (editor), *Encyclopedia of Nonlinear Science*, (New York:Taylor and Francis) 2004. 218
17. C. Baesens, *Variational problems in solid state physics*, (unpublished) Lecture Notes (University of Warwick, 1994). 219
18. A. Katok and B Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems* (Cambridge: CUP, 1995). 219
19. C. Golé, *Symplectic Twist Maps: Global Variational Techniques* (Advanced Series in Nonlinear Dynamics vol 18) (Singapore: World Scientific, 2001) 219, 220
20. A.A. Middleton, *Phys. Rev. Lett.* **68**, 670 (1992). 220
21. L.M. Floría and J.J. Mazo, *Adv. Phys.* **45**, 505 (1996). 220, 221
22. C. Baesens and R.S. Mackay, *Nonlinearity* **11**, 949 (1998). 220, 221
23. C. Baesens and R.S. Mackay, *Nonlinearity* **17**, 949 (2004). 221
24. L. Kuipers and H. Niederreiter *Uniform distribution of sequences* (John Wiley and Sons, New York 1974). 222
25. *Appl. Phys A* **75** (2002), special issue on "Ratchets and Brownian motors", guest ed. H. Linke. 223
26. L.H. Tang, *Ph.D. Thesis*, (Carnegie-Mellon University 1987); L.H. Tang and R.B. Griffiths *J. Stat. Phys.* **53**, 853 (1988). 225
27. L.M. Floría, F. Falo, P.J. Martínez and J.J. Mazo, *Europhys. Lett.* **60**, 174 (2002). 228, 233
28. L.M. Floría, P.J. Martínez, S. Flach and M.V. Fistul, *Physica D*, **187**, 100 (2004). 230

Riddled Basins and Coupled Dynamical Systems

P. Ashwin

Department of Mathematical Sciences, Laver Building, University of Exeter, Exeter EX4 4QE, UK
P.Ashwin@ex.ac.uk

1 Introduction

The collective dynamics of groups of coupled dynamical systems is of great interest for understanding spontaneous pattern formation in biological and many other systems; see for example [58]. One can learn a lot about such systems by first studying idealized cases where the systems are perfectly identical; this approach has been very successful in understanding general properties of synchronization as well as particular applications; see for example [77]. In this chapter we consider how this can lead to the appearance of attractors with riddled basins. These basins appear because symmetries of dynamical systems force the presence of invariant submanifolds; the attractors within invariant manifolds may be only weakly attracting transverse to the invariant manifold and this leads to a basin structure that is, roughly speaking, full of holes.

From a theoretical point of view, this behaviour is of interest because it seems strange or pathological but is in some sense common. From a practical point of view, this behaviour points towards the presence of extreme sensitivity of the dynamics to noise, also called 'bubbling' of attractors. Most interestingly, if we consider generic dynamics within a class of symmetric systems, riddled basins can appear as a robust phenomenon; they can be persistent for open sets of parameters of the system.

For the remainder of this section we briefly discuss basins of attraction and a motivating example of a piecewise linear map with an explicitly computable riddled basin attractor. More general properties of riddled sets and basins are discussed in Sect. 2 including their noise sensitivity. This is followed in Sect. 3 by a discussion of the use of symmetries, ergodic measures and Lyapunov exponents tools for identifying riddled basins; we also discuss anisotropic riddling in Sect. 4 along the lines of [7]. Finally in Sect. 5 we outline a few open problems related to riddling phenomena.

P. Ashwin: *Riddled Basins and Coupled Dynamical Systems*, Lect. Notes Phys. **671**, 181–207 (2005)
www.springerlink.com

1.1 Attractors for Smooth Maps on Compact Manifolds

Qualitative dynamics is about ignoring information. Since one cannot find explicit solutions to all but the simplest dynamical systems, the first step is usually to concentrate on what happens "eventually", i.e. for the asymptotic dynamics, and this leads to several possible definitions of attractor. We will discuss a couple of these notions, though there are other notions of attractor based on invariant measures are also very useful; see for example [29, 71]. We refer the reader to the other chapters in this volume for further discussion and examples of attractors in coupled systems as well as [61].

We concentrate on dynamical systems generated by iterated maps

$$f : M \to M$$

where M is a compact manifold and f is a smooth map generating a dynamical system on iteration, though in some cases we may drop assumptions of smoothness or may wish to consider flows F_t with $t \in \mathbb{R}$. There are some comments on generalizations of this to other less restrictive cases in later sections, in particular in Sect. 5.

The (forward) trajectory through $x \in M$ is the set $\{f^n(x) \ : \ n = 0, 1, 2, \cdots\}$ and we say a set $A \subset M$ is (forward) invariant if $f(A) = A$. Define the ω limit set by

$$\omega(x) = \bigcap_{n>0} \overline{\{f^m(x) \ : \ m > n\}} \, .$$

This represents the set of points that the orbit of x accumulates on as $t \to \infty$.

Note that the set $\omega(x)$ is invariant under f. To see this, consider $y \in \omega(x)$; then there is a sequence n_k such that $f^{n_k}(x) \to y$. Continuity of f means that

$$f^{n_k+1}(x) \to f(y)$$

and so $f(y)$ is also in $\omega(x)$. Hence $\omega(x)$ is (forward) invariant.

In cases where f is invertible one can apply the same considerations to $\alpha(x)$, the limits of $f^{-n}(x)$ as $n \to \infty$. An attractor is in some sense the smallest set that contains all $\omega(x)$ limits that one care about; since ω-limit sets are invariant we only really need to consider invariant sets as candidates for attractors. Given an invariant set A, consider the set of points whose orbits are asymptotic to A

$$\mathcal{B}(A) = \{x \in M \ : \ \omega(x) \subset A\}$$

which is the *basin of attraction* of A.

For the definition of attractor given by Milnor [56] we need a Lebesgue equivalent background measure on M that we denote by $\ell(\cdot)$ and by $\ell(A) > 0$ we include the possibility that it is infinite. Recall that Lebesgue measure is simply a generalization of length/area/volume such that one can measure

many sets constructed from infinite unions and intersections of open sets. We say the compact invariant set A is a *weak attractor* if (M1) holds, it is a *Milnor attractor* if (M1) and (M2) hold, and a *minimal Milnor attractor* if (M1) and (M2') hold [16], where:

(M1) $\ell(\mathcal{B}(A)) > 0$.
(M2) For any proper compact invariant subset $A' \subset A$ we have

$$\ell(\mathcal{B}(A) \setminus \mathcal{B}(A')) > 0.$$

(M2') For any proper compact invariant subset $A' \subset A$ we have

$$\ell(\mathcal{B}(A')) = 0.$$

for further comments on these notions see also [24].

One can think of a Milnor attractor as the smallest compact set that attracts all initial conditions except for a set of zero measure with respect to some natural "background measure". As shown in [56] Milnor attractors can be constructed by examining the likely limit sets for positive measure subsets $S \subset M$; these are compact sets $\Lambda(S)$ that are the smallest such that $\omega(x) \subset \Lambda$ except for a zero measure set of $x \in S$.

From the point of view of numerical simulations of a dynamical system, Milnor attractors are just as reasonable a definition of attractor as for example an asymptotically stable attractor; recall that an *asymptotically stable attractor* is an A such that (i) for any open set U containing A there is an open set V containing A such that $x \in V$ implies that $f^n(x) \in U$ for all n and (ii) $\omega(x) \subset A$ for all $x \in V$. An A that satisfies only (i) is Lyapunov stable. Note that an asymptotically stable attractor must be a weak Milnor attractor with a basin than contains an open set.

1.2 A Motivating Example

Simple examples of systems with riddled basin attractors can be found by considering skew product dynamical systems. A *direct product* of two maps g and h is simply the map obtained by $f(x, y) = (g(x), h(y))$; g and h are referred to as factors; projecting onto one of the coordinates gives a map that is well-defined. A *skew product* is a map of the form

$$(x, y) \mapsto f(x, y) = (g(x), h(x, y)) \tag{1}$$

that has only one factor, in this case g. In cases where $h(x, y) = y\hat{h}(x, y)$ the map (1) has an invariant subspace $N = \{(x, 0)\}$. We consider as in [6] the piecewise linear map on $(x, y) \in [0, 1] \times [0, \infty)$ of the form (1) where

$$g(x) = \begin{cases} \alpha^{-1}x & \text{for } 0 \le x < \alpha \\ (1-\alpha)^{-1}(x-\alpha) & \text{for } \alpha \le x < 1 \end{cases} \tag{2}$$

$$h(x, y) = \begin{cases} \gamma y & \text{for } y < 1 \text{ and } 0 \le x < \alpha \\ \gamma^{-1} y & \text{for } y < 1 \text{ and } \alpha \le x \le 1 \\ 1 + \beta(1-y) & \text{for } y \ge 1\ . \end{cases} \tag{3}$$

We will assume that $\gamma > 1$, $0 < \alpha < \frac{1}{2}$ and $-1 < \beta < 0$ are fixed. Varying α through $1/2$ allows one to observe a blowout bifurcation of an attractor $A = N$. Varying β allows one to change the dynamics between supercritical and subcritical scenarios; for more details of the other parameter values, see [6]. In the case $-1 < \beta < 0$, any trajectory that arrives in $y \leq 1$ will stay there for ever.

First, observe that this map has an invariant subspace $y = 0$ on which the dynamics is chaotic on this set A in the strongest sense that is commonly in use; the dynamics has Lebesgue measure as an ergodic invariant measure for which the dynamics is Bernoulli.

Now pick any $x \in [0, 1]$ and consider its itinerary under the "skewed doubling map" g. This defines a symbol sequence $\{s_i\}_{i=0,1,2\ldots}$ where $s_i = 0$ if $0 \leq f^i(x) < \alpha$ and $s_i = 1$ if $\alpha \leq f^i(x) \leq 1$. Define

$$l_k = \#\{0 \leq j < k \ : \ s_j = 0\}, \ r_k = \#\{0 \leq j < k \ : \ s_j = 1\},$$

the number of times that the itinerary of x is resp. to the left/right of $x = \alpha$. For almost all x we can use the fact that Lebesgue measure on $[0, 1]$ is invariant and ergodic under g to conclude that the following sequences converge

$$\lim_{k\to\infty} \frac{l_k}{k} = \alpha, \ \lim_{k\to\infty} \frac{r_k}{k} = 1 - \alpha \tag{4}$$

for almost all $x \in [0, 1]$. If we now define

$$M_k(x) = \gamma^{l_k - r_k}$$

then as long as $\alpha < \frac{1}{2}$ we have by (4) that $\lim_{k\to\infty} \frac{1}{k}(l_k - r_k) = 2\alpha - 1 < 0$ so that $\lim_{k\to\infty} l_k - r_k = -\infty$. Hence

$$\lim_{k\to\infty} M_k(x) = \lim_{k\to\infty} \exp\left[(l_k - r_k)\ln\gamma\right] = 0$$

for almost all x. By comparing with (3) one can verify that

$$f^k(x, y) = (g^k(x), M_k(x)y) \tag{5}$$

as long as $M_k(x)y$ does not exceed 1. We define

$$Y(x) = \max\left(1, (\sup_{k\geq 0} M_k(x))^{-1}\right)$$

For almost all x we have $0 < Y(x) \leq 1$. If $y < Y(x)$ then (5) holds for all $k \geq 0$ because $M_k(x)y$ will never exceed 1.

In summary, there is a function $Y(x)$ with $Y(x) > 0$ for almost all x that describes the basin of attraction of A:

$$\mathcal{B}(A) = \{(x, y) \ : \ 0 \leq y < Y(x)\}.$$

Clearly, the Lebesgue measure of $\mathcal{B}(A)$ must be greater than zero; it is simply $\ell(\mathcal{B}(A)) = \int_{[0,1]} Y(x)\,dx$, and A is a minimal Milnor attractor because almost all x have g orbits that are dense in $[0,1]$.[1] However the function $Y(x)$ is highly non-smooth and it is this that makes the basin $\mathcal{B}(A)$ in fact riddled as shown in Fig. 1.

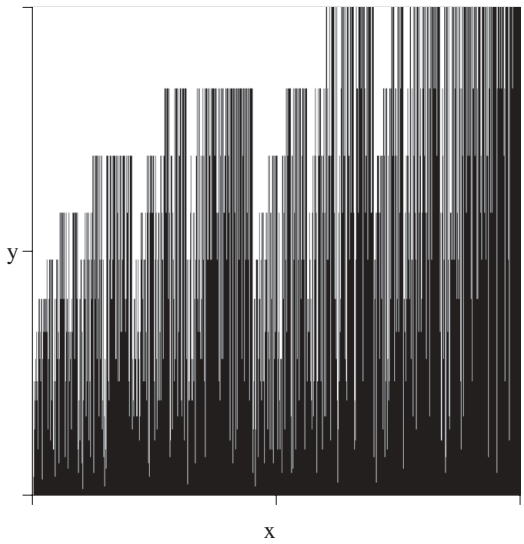


Fig. 1. The *black* set shows a numerical approximation of the riddled basin for the attractor in $(x, 0)$ with $\alpha = 0.45$, $\gamma = 1.2$ and $\beta = -1$ for the map (1,2,3). The *box* shows points $(x, y) \in [0,1]^2$ while initial conditions in the white set are ejected to $y \geq 1$

For this example we can compute the measure within the basin as in [6]. Let $\epsilon = 2\alpha - 1$ and note that $\alpha < 1/2$ implies that $\epsilon < 0$ in what follows. We partition $[0,1]^2$ into a set of strips

$$I_n = [0,1] \times (\gamma^{-n-1}, \gamma^{-n}),\ \ n = 0, 1, 2, \ldots$$

where the strip I_n has height $\gamma^{-n-1}(\gamma - 1)$. The form of the map means that it is conjugate to a mapping on the I_n defined by

$$T(x, n) = (g(x), m(x, n))$$

where

$$m(n, x) = \begin{cases} n - 1 & \text{if } 0 \leq x < \alpha \text{ and } n > 0 \\ n + 1 & \text{if } \alpha \leq x \leq 1 \text{ and } n > 0 \\ 0 & \text{if } n = 0 \end{cases}$$

[1] Recall that a sequence of points x_n is *dense* in a metric space M if any open set in M contains a point in the sequence.

which can be viewed as a biased random walk on $\mathbb{N} \cup \{0\}$ with 0 being an absorbing state. Let q_n, $n > 0$, be the probability of arriving at the state 0; then

$$q_n = \frac{1}{2}(1+\epsilon)q_{n-1} + \frac{1}{2}(1-\epsilon)q_{n+1}$$

which has solution $q_n = \Lambda^{-n}$ where $\Lambda = (1-\epsilon)(1+\epsilon)$ will satisfy $\Lambda > 1$. Hence the measure of points that are attracted to the invariant subspace is given by

$$\ell(\mathcal{B}(A)) = \sum_{n=1}^{\infty} \gamma^{-n-1}(\gamma-1)(1-\Lambda^{-n}) = \frac{\gamma(\Lambda-1)}{\Lambda\gamma-1} \cdot$$

Clearly $\ell(\mathcal{B}(A)) \to 0$ as $\epsilon \to 0^-$. We illustrate in Fig. 2 the structure of the complement of the basin of attraction; note that there are "tongues" of instability that come down to touch the x-axis at all points $(x, 0)$ such that the itinerary of x ends in an infinite number of 0s. Since this set is dense, we can conclude that the basin of attraction of $y \geq 1$ is dense and all points must exit after a finite time, it follows from this that this set is open and dense. In terms of $Y(x)$ this means that $Y(x) = 0$ on a dense set in $[0, 1]$; since $Y(x)$ is almost everywhere positive, $Y(x)$ is discontinuous on a dense set on $[0, 1]$. In fact, $Y(x)$ is upper semicontinuous at almost all x and the set $\{(x, y) \ : \ y > Y(x)\}$ is open and dense in $\mathbb{R}^2$.

One might suspect that this behaviour is caused by the presence of discontinuities in the map (1) but in fact this is not the case; similar basins appear

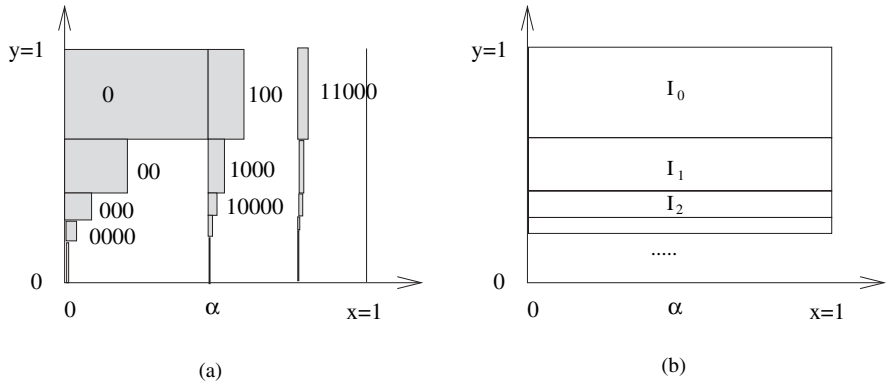


Fig. 2. (**a**) The structure of the basin of attraction of A for the map (1,2,3) discussed in the text. For $\alpha < 0$ the invariant set A in $y = 0$ has a basin with positive measure whose complement is open and dense and whose structure is shown by the *shaded* set. The coding indicates the x-itinerary of those points in $y < 1$ before they are expelled to $y \geq 1$. By including all possible finite words that occur before expulsion one obtains a set that is dense and open in $[0, 1]^2$ yet which does not have full measure. (**b**) shows the strips I_n for this map (see text)

naturally in smooth and even invertible maps and seem to be robust in many systems with symmetries or invariant subspaces.

1.3 Related Notions of Basin Complexity

A somewhat simpler concept than riddled basin, is that of a *fractal basin boundary*. It has been observed since the work of Julia and Fatou for complex maps, that even linearly stable equilibria e may have basins of attraction $B = \mathcal{B}(\{e\})$ with boundary such that the Hausdorff dimension $\dim_H(\partial B)$ is not an integer; for example, see [32, chapter 14] (recall that $\partial B = \overline{B} \setminus \text{Interior}(B)$). One should stress however that for riddled basins, not just the boundary has "fractal" properties, but the whole set is inseparable from its boundary and in fact $\partial B = B$ up to a set of zero measure.

Another notion of attraction that is weaker than asymptotic stability but is stronger than Milnor attraction is that of *essential asymptotic stability* [54]; we say an invariant set A is e.a.s. if it is asymptotically stable if one excludes a set of small measure compared to small neighbourhoods of A; more precisely if there is a set $\mathcal{S}$ such that for any neighbourhood U of A and any $0 < a < 1$ there is a neighbourhood V of A with $\ell(V \setminus S)/\ell(V) > a$ where $x \in V \setminus \mathcal{S}$ implies that $f^n(x)$ remains in U and is asymptotic to A. Such attractors are found quite commonly and robustly in heteroclinic networks that have lost asymptotic stability. However in this case the basin of attraction of A may still be an open unriddled set.

2 Riddled Sets and Riddled Basins

In contrast to attractors with fractal basin boundaries, a riddled basin is "fractal" everywhere. However, since a basin of attraction must have positive measure within phase space, in fact it must have Hausdorff dimension equal to that of phase space and so cannot have non-integer dimension. It is a "fat fractal" in the terminology of [30, 34] in that it contains a dense set of holes. In the following we will use the definitions as in [16] though we note that there are several possible equivalent definitions. We denote by $B_\delta(x) = \{y \in M \ : \ |y - x| < \delta\}$ the open δ-ball about x in M.

A *riddled* subset $A \subset \mathbb{R}^n$ is a measurable set with the property that for any $\delta > 0$ and any $x \in A$ we have

$$\ell(A \cap B_\delta(x)) > 0 \quad \text{and } \ell(A^c \cap B_\delta(x)) > 0 \, . \tag{6}$$

More generally given any set A with positive measure one can define its *riddled component*

$$A_{rid} = \{x \in A \ : \ \ell(B_\delta(x) \cap A)\ell(B_\delta(x) \cap A^c) > 0 \quad \text{for all } \delta > 0\} \, .$$

We use this to distinguish cases of full riddling ($\ell(A \setminus A_{rid}) = 0$), lack of riddling ($\ell(A_{rid}) = 0$), with partial riddling being any other case. If A is the closure of an open set with smooth boundary then $\partial A = A_{rid}$.

Sometimes one would like to discuss a similar condition to (6) but rather than comparing A and its complement A^c one would like to compare to disjoint subsets A and B of $\mathbb{R}^n$. If A and B are disjoint and both have positive measure then we say A is *riddled with* B in the case that almost all $x \in A$ and all $\delta > 0$ have

$$\ell(B_\delta(x) \cap A)\ell(B_\delta(x) \cap B) > 0 \, .$$

In the case that A is riddled with B and B is riddled with A we say they are *intermingled*.[2]

One can show that the riddled component of an invariant set A is invariant for a large class of maps. We say f is of type (P) if it is continuous, a local homeomorphism and nonsingular (i.e. for any V, $\ell(V) = 0$ if and only if $\ell(f(V)) = 0$). The following result taken from [16] can be generalised to cases where f is almost everywhere a local homeomorphism or to some classes where f is invertible but discontinuous on a set of zero measure [12].

Theorem 2.1. *Suppose that $f : M \to M$ is of type (P) and V is invariant, then V_{rid} is invariant.*

Proof. Consider $x \in V$ and choose a neighbourhood U_1 of x and U_2 of $f(x)$ such that $f : U_1 \to U_2$ is a homeomorphism. Consider any $\delta > 0$ such that $B_\delta(x) \subset U_1$ and $B_\delta(f(x)) \subset U_2$. Continuity of f means that we can find $\delta > \epsilon > 0$ such that $f(B_\epsilon(x)) \subset B_\delta(f(x))$. Hence $\ell(B_\delta(f(x)) \cap V^c) \geq \ell(f(B_\epsilon(x)) \cap V^c)$; by considering the local inverse we see that there is a $\delta > 0$ such that $\ell(B_\delta(x) \cap V^c) = 0$ if and only if there is an $\epsilon > 0$ such that $\ell(B_\epsilon(f(x)) \cap V^c) = 0$. Applying the same argument to V in place of V^c means that x is in V_{rid} if and only if $f(x)$ is. □

This result can adapted to the case where f is almost everywhere a local homeomorphism in which case we conclude that V_{rid} is invariant up to a set of zero measure.

2.1 Characterization of Riddled Basins

One approach to characterizing dynamical invariants of riddled basins has been to compute their *uncertainly exponent*; [34, 62]. This allows one to compute a quantity that characterizes the "riddledness" of the basin or more generally of a "fat fractal". We define this as in [62] for a riddled subset of $\mathbb{R}^2$.

[2] If two positive measure sets A and B are intermingled then they cannot be evenly distributed in the following sense; the Lebesgue density theorem [82, p107] or [32, p69] implies that almost all points in A are points of density for A in the sense that for almost all $x \in A$, $\lim_{\epsilon \to 0} \ell(B_\epsilon(x) \cap A)/\ell(B_\epsilon(x)) = 1$. Hence almost all points in $A \cup B$ are density points for only one of the sets A or B.

Pick a typical line transverse to the basin and choose two points within 2ϵ of each other. If we can estimate the probability p that one point is in the basin and the other is not by $p \sim \epsilon^{\phi}$ we say that there is an uncertainty exponent ϕ; in [62] it is shown that one can estimate this as the ratio $\phi = (\lambda^{\perp})^2/4D\lambda^{||}$ where $\lambda^{\perp}$ and $\lambda^{||}$ are the transverse and tangential Lyapunov exponents and D is the rate of convergence of variance of finite time Lyapunov exponents.

Another (simpler) characterization of a riddled basin proposed by [62] is the scaling of the measure near the attractor; this corresponds to taking a line at a distance y from the attractor; in many cases the measure of the basin intersected with this line scales as y^{η}. This is estimated in [62] as being $\eta \sim |\lambda^{\perp}|/D$.

An interesting observation is that the Kaplan-Yorke formula relating dimension of attractor to Lyapunov dimension [44] can fail for riddled basin attractors; one characteristic of blowout bifurcations is that they are associated with a sudden regain of the validity of that formula.

2.2 Locally and Globally Riddled Basins

There are cases where an attractor has a basin with full measure in some open neighbourhood but still a form of riddling in the convergence towards the attractor. More precisely, consider any open neighbourhood U of A and define the *basin of A relative to U* to be

$$\mathcal{B}_U(A) = \{x \in M \ : \ \omega(x) \subset A \quad \text{and } f^n(x) \subset U \quad \text{for all } n \geq 0\} \ .$$

We say the basin of A is *locally riddled* if there is a neighbourhood U of A such that $\mathcal{B}_U(A)$ is riddled [16] (this is a stronger assumption than that given in [9] where it is only assumed that the riddled component of $\mathcal{B}_U(A)$ is dense in A). It is *globally riddled* in the case that U can be chosen to be equal to M.

2.3 Riddling and Noise Sensitivity; Bubbling of Attractors

A useful model for noise in iterated maps consists of adding an independent uniformly distributed random variables to all components at every iteration. For attractors with riddled basins, this can give rise to discontinuous behaviour in the support of attractors as the noise goes to zero and was called *bubbling* in [8].

More precisely, suppose that $A \subset N \subset M$ with A a Milnor attractor for $f : M \to M$ and the basin of A is locally riddled (say $\mathcal{B}_U(A)$ is riddled with U compact in M) but which is asymptotically stable in the invariant subspace N. Consider the perturbed map

$$x_{n+1} = f_{\sigma}(x) = f(x_n) + \sigma\xi_n \tag{7}$$

where ξ_n is a vector of i.i.d. random variables uniformly distributed in $[-0.5, 0.5]$. Suppose that f_{σ} has an attractor A_{σ} in the sense that A_{σ} is the

smallest compact set that contains all limit points for realizations of the noise ξ_n and all x in some open region.

We say the attractor A for f is *stable to noise* if $A_\sigma \to A$ in the Hausdorff metric as $\sigma \to 0$. This is a weaker notion than for example stochastic stability discussed in [18]; the latter considers convergence of measures for the perturbed system to a natural measure for the noise-free system. Define the *unstable set* of A to be

$$\mathcal{U}(A) = \{x \ : \ \exists \{x_{-n}\} \quad \text{with } f(x_{n-1}) = x_{-n} \ \& \ \lim_{n\to\infty} d(x_{-n}, A) = 0\}$$

i.e. the set of points that have a backward trajectory from A (this definition is simpler if f is invertible); clearly $\mathcal{U}(A) \supset A$. If A is stable to noise then $\mathcal{U}(A) \subset A$. Riddled basin attractors typically have $\mathcal{U}(A) \not\subset A$ and so are not stable to noise.

We will explore further properties of bubbling (in the presence of several invariant subspaces) in Sect. 4.

3 Symmetry, Transverse Stability and Riddling

Consider a smooth iterated mapping $f : M \to M$ with $M = \mathbb{R}^m$ some Euclidean space and write $\ell_N(.)$ to denote Lebesgue measure on N a linear subspace of M. Suppose that f commutes with (is equivariant for) the action of some finite matrix group Γ acting on M (we only consider finite groups here though similar results can be obtained for compact groups). This means that for any $x \in M$ and for any $g \in \Gamma$ we have that

$$g.f(x) = f(g.x) \ .$$

In general this will force a number of linear subspaces of M to be f-invariant. More precisely, given any subgroup Σ of Γ we define the *fixed point subspace*

$$\mathrm{Fix}(\Sigma) = \{x \in M \ : gx = x \quad \text{for all } g \in \Sigma\}$$

which is f-invariant. Not all subgroups give rise to distinct fixed point subspaces; those that do are the *isotropy subgroups* of the action of Γ on M; these are the subgroups $\Sigma(x)$ with $x \in M$ such that $\Sigma(x) = \{g \in \Gamma \ : \ gx = x\}$; these are precisely the possible symmetries of points in M. We therefore obtain, for a given group action, a finite number of linear subspaces $N_i \subset M$, $i = 1 \cdots n$ and isotropy subgroups $\Sigma_i \leq \Gamma$ such that $N_i = \mathrm{Fix}(\Sigma_i)$ and so $f(N_i) \subset N_i$. Note that $N_i \subset N_j$ if and only if $\Sigma_i \geq \Sigma_j$. Observe also that $N_i \cup N_j = N_k$ for some k is also invariant. Conversely, if Σ_k is the smallest isotropy subgroup that contains both Σ_i and Σ_j then $N_i \cup N_j = N_k$.

Two subgroups Σ_1, Σ_2 are *conjugate* if there is a $g \in \Gamma$ such that $\Sigma_1 = g^{-1}\Sigma_2 g$. The next basic result states that conjugate subgroups have fixed point spaces that are mapped onto each other by the group. More precisely we have the following elementary result.

Lemma 3.1. *Suppose that Σ is an isotropy subgroup and $g \in \Gamma$. Then $g^{-1}\Sigma g$ is an isotropy subgroup and* $\mathrm{Fix}(g^{-1}\Sigma g) = g^{-1}\mathrm{Fix}(\Sigma)$.

Proof. Note that

$$\begin{aligned}\mathrm{Fix}(g^{-1}\Sigma g) &= \{x \in M \ : \ g^{-1}hgx = x \quad \text{for all } h \in \Sigma\} \\ &= g^{-1}\{gx \in M \ : \ hgx = gx \quad \text{for all } h \in \Sigma\} = g^{-1}\mathrm{Fix}(\Sigma)\ .\end{aligned}$$

□

There has been some work on symmetries of general chaotic attractors in symmetric systems; see [33] for an overview, and [17, 19, 25, 27, 55]. Other phenomena that can appear include intermittent dynamics between states with a variety of different symmetries; for example [13, 14, 31].

3.1 Example; Invariant Subspaces for Four Globally Coupled Maps

Consider the map on $\mathbb{R}^4$ defined by

$$x_i' = (1-\epsilon)f(x_i) + \frac{\epsilon}{4}\sum_{j=1}^{4} f(x_j) + \sigma_i\eta_i \tag{8}$$

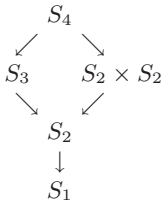
for $i = 1, \ldots, 4$, where at each time-step each η_j is an independent random variable that is uniformly distributed on $[-0.5, 0.5]$. The local map $f(x) = 1 - ax^2$ is a quadratic map and ϵ is the coupling strength. The noise in the ith component can be controlled by setting σ_i nonzero.

The system (8) has been studied by several authors including notably Kaneko [42] although only with isotropic noise perturbations, and displays a wide range of synchronization and chaotic behaviour. Taborev et al. [79] have also recently looked at the noise-free case of $n = 3$ cells in some detail.

This map has the symmetry of all permutations on n objects; it is equivariant under the action of the group S_4 given by permutation matrices on $\mathbb{R}^4$. There are many invariant subspaces corresponding to isotropy subgroups; these can be characterized by partitions of $\{1..4\}$ into $1 \le m \le 4$ groups of identical cells and after identifying conjugate subgroups the possible isotropy subgroups are conjugate to one of the *partial clustering states* S_1, S_2, $S_2 \times S_2$, S_3 and S_4. These have fixed point subspaces that are given by coordinates within one partition being equal; for example if H is the subgroup generated by the two-cycles (12) and (34) then $\mathrm{Fix}(H) = \{(x, x, y, y)\}$.

One can order the possible subspaces by $\Sigma_1 < \Sigma_2$ if and only if $\mathrm{Fix}(\Sigma_2) \subset \mathrm{Fix}(\Sigma_1)$; for the case $\Gamma = S_4$ this gives the containments as shown in the *isotropy lattice* in Table 1. Note that for n larger the isotropy lattice of S_n becomes much more complicated.

Table 1. The lattice of isotropy subgroups up to conjugacy for the symmetry group S_4 of four globally coupled maps. The arrows indicate containment of subgroups; the isotropy subgroups can be interpreted as cluster states for the coupled maps



3.2 Symmetries and Lyapunov Exponents

Now suppose we have a compact invariant set $A \subset N \subset M$ with N a linear invariant subspace. Assume that A supports an ergodic f-invariant probability measure μ_{nat} that is a *natural measure*, i.e. there is a positive measure set (with respect to Lebesgue measure on N) such that points in these sets have ergodic averages determined by μ; i.e. such that

$$\lim_{k\to\infty} \frac{1}{k} \sum_{j=0}^{k-1} \phi(f^j(x)) = \int_A \phi(y)\, d\mu(y)$$

for any continuous $\phi : N \to \mathbb{R}$ and a positive measure set of $x \in N$. We also define

$$\mathrm{Erg}(A) = \{\mu \ : \quad \text{ergodic measures with support on } A\}.$$

With respect to any ergodic measure $\mu \in \mathrm{Erg}(A)$ we define the *Lyapunov exponents* (L.E.s)

$$\lambda(x, v) = \lim_{n\to\infty} \frac{1}{n} \log \frac{|Df^n(x)v|}{|v|} \ .$$

Oseledec's theorem [29, 53, 60, 66, 71] implies that for μ-almost all x and any v this limit converges and may take one of only a finite number of possible values $\lambda_i(\mu)$, $i = 1, \ldots, m$ (we count multiplicity by the dimension of the subspace of v that give this value for typical x). Recall that given any ergodic measure with support contained within an invariant subspace N, the Lyapunov exponents (L.E.s) will split into two groups [9, 23]; the *tangential L.E.s* $\lambda_j^{||}(\mu)$ and the *transverse L.E.s* $\lambda_j^{\perp}(\mu)$; the former correspond to perturbations v within N (cf the Sacker-Sell spectrum [75]). We assume that for any ergodic μ the transverse L.E.s are ordered greatest first; $\lambda_1^{\perp}(\mu) \geq \lambda_2^{\perp}$ etc. Define

$$\Lambda_{\max} = \sup_{\mu\in\mathrm{Erg}(A)} \lambda_1^{\perp}(\mu), \ \Lambda_{nat} = \lambda_1^{\perp}(\mu_{nat}).$$

In the case that $\Lambda_{\max} > 0$ and $\Lambda_{nat} < 0$ we will obtain an attractor A whose basin is locally riddled and possibly globally riddled, and at least there will be

a neighbourhood U for which the attractor have a relative basin $\mathcal{B}_U(A)$ whose riddled component includes A. It is however more difficult to come up with necessary conditions for globally riddled basins that are not very restrictive. See [7, 9, 16, 24] for some conditions that imply local riddling.

We say a Milnor attractor is *regular* if for any open neighbourhood U of A then the measure of the basin of A relative to U is positive; $\ell_M(\mathcal{B}_U(A)) > 0$. A result in [1] implies that any Milnor attractor that is a uniformly hyperbolic within N with $\Lambda_{nat} < 0$ will be regular.

Note that for uniformly hyperbolic A (within N) all ergodic measures on A are limit points of sequences of periodic measures supported within A [78] and so for such sets the existence of a measure with positive transverse Lyapunov exponent implies the existence of periodic points with L.E.s that are arbitrarily close to that of the natural measure.

3.3 Examples of Riddled and Intermingled Basins

Since the phenomena of riddled basins was uncovered [1], riddled basins have been found in a range of applications, for example learning dynamical systems [59], coupled chaotic oscillators [8], mechanical systems [83] and electronic systems [38] and especially coupled maps, e.g. [36] where the literature is too extensive to list; see other chapters in this book.

As an example, we consider here a smooth map from [11] with a riddled basin attractor:

$$f(x, y) = (rx(1 - x) + sxy^2, \nu e^{-x} y + y^3) \tag{9}$$

with $r = 4$, $s = 0.3$ and $\nu = 1.5$. This has an attractor within $y = 0$ given by $A = [0, 1] \times \{0\}$ on which the dynamics is that of a logistic map with $r = 4$.[3] Evaluating the transverse Lyapunov exponent with respect to any ergodic measure μ within A amounts to computing the integral

$$\lambda^{\perp}(\mu) = \int (-x + \ln(\nu))\, d\mu(x) = \ln \nu - \int x\, d\mu(x).$$

Observe that (9) has a fixed point in A at $(0, 0)$ has the largest possible $\lambda^{\perp} = 0.40546511$ meaning that $\Lambda_{\max} = 0.40546511$, while the natural measure gives $\Lambda_{nat} = -0.094534$. We can use the topological conjugacy of the map f on A to a doubling map to conclude that there is a dense set of preimages of the fixed point at $(0, 0)$; hence we expect the basin of attraction of A to be locally riddled. In this case we can observe that it is globally riddled; see Fig. 3. Note that unlike the example (1) this map is neither discontinuous nor a skew product, as long as $s \neq 0$.

Examples of intermingled basins have been found in maps [2] (three intermingled basins) and [41] as well as coupled ODEs [28]. We present an example

[3] In this case the basin of A within $y = 0$ contains no points outside of A.

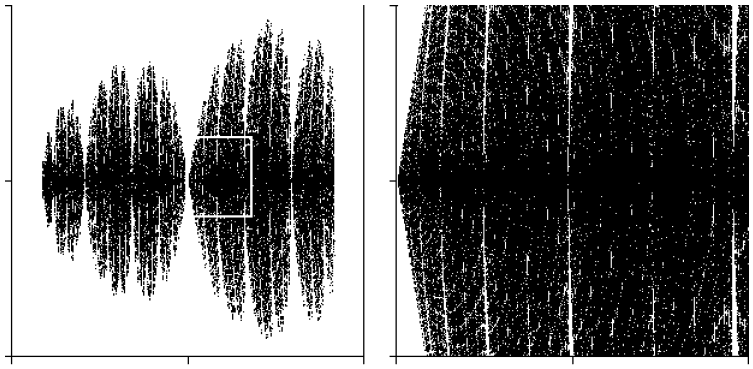


Fig. 3. Numerical approximation of the riddled basin of attraction of an attractor in $y = 0$ for the smooth map (9) with $r = 4$, $s = 0.3$ and $\nu = 1.5$. The basin is shown in black, while the points in white converge to attractors at infinity. The image (**a**) shows the basin in the area $[-0.1, 1.1] \times [-0.6, 0.6]$ while (**b**) shows a zoom into $[0.5, 0.7] \times [-0.1, 0.1]$

in a map due to Ding and Yang [28] of two intermingled basins for a map on $(x, y) \in [-1, 1]^2$ given by $f(x, y) = (x', y')$ where

$$\begin{aligned} x' &= g(x) + \epsilon(g(y) - g(x) + g(y)^3 - g(x)^3) \\ y' &= g(y) + \epsilon(g(x) - g(y) + g(x)^3 - g(y)^3) \end{aligned} \tag{10}$$

for $g(x) = 3.4x(1 - x^2)e^{-x^2}$ and $\epsilon = 0.48$. This map has attractors $A^{\pm}$ in the diagonal $x = y$ on either side of the origin. Figure 4 illustrates the basin of the attractor A^+ in black; the approximation is obtained by computing

$$A_n = f^{-n}([0, 1] \times [-1, 1]).$$

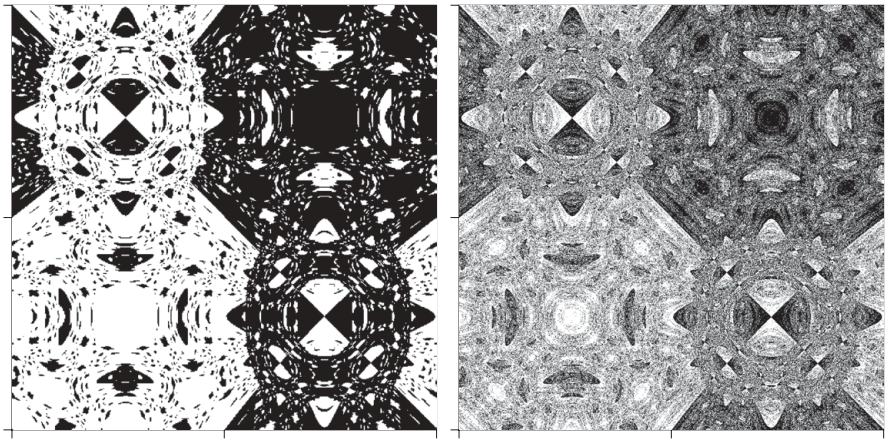
If we assume that $A^{\pm}$ are the only Milnor attractors for f then

$$\bigcap_{m>n} \left(\bigcup_{p>m} A_p \right)$$

will converge (apart from a set of zero measure) to $\mathcal{B}(A^+)$ as $n \to \infty$. The sets A_6 and A_{25} are shown; the Lebesgue density theorem implies that almost all points are points of density for either the white or the black set (though there are clearly do exist points such as $(0, 0)$ that are points of density for neither white nor black set; it is even possible that these are dense in $[-1, 1]^2$).

3.4 Normal and Non-normal Parameters

Most analytical studies of blowout bifurcation rely on being able to vary the dynamics while preserving the dynamics on some invariant subspace. This



(a) (b)

Fig. 4. Numerical approximation of intermingled basins of attraction of an attractor A^+ in $x = y$, $x > 0$ for the map (10) with $\epsilon = 0.48$. The basin is shown in black, while the points in white are an approximation of the set basin of A^-. Both images show the area $(x, y) \in [-1, 1]^2$; (**a**) shows an approximation using 6 iterates while (**b**) uses 25 iterates

allows one to effectively vary the transverse Lyapunov exponents for fixed invariant measures. More precisely, if $f_r : M \to M$ is parameterized smoothly by $r \in \mathbb{R}$ and if $f_r(N) = N$ for all r and some fixed submanifold then we say r is a *normal parameter* if $f_r|_N$ is independent of r. Otherwise we refer to r as a *non-normal parameter*. For normal parameters the Lyapunov exponents will vary with the parameter in a way that may or may not be analytic; see for example [9, 73, 74]. In cases that the normal Lyapunov exponent does vary continuously, the transition to riddling of a basin can be determined by finding the point at which it loses asymptotic stability.

As discussed in [11, 26] the transitions obtained on varying a non-normal parameter will typically be much more complicated than on varying a normal parameter, unless the attractor within N is robust [20]. We refer to the other chapters in this volume for discussion of the blowout bifurcation.

4 Anisotropic Riddling in Coupled System

Following [7] we consider properties of riddling that can appear in the presence of more than one invariant subspace. Riddling of the basin may occur in some directions but not others; we say that riddling is typically *anisotropic* in transverse directions unless they are symmetrically related in ways that we characterize in Theorems 4.1, 4.2. In more general cases the dynamics can display anisotropic sensitivity to noise. We also illustrate internal riddling transitions where the number of directions in which riddling occurs changes.

For nested invariant subspaces N and P with $A \subset N \subset P \subset M$ we define

$$\{\lambda_i^{N,P}(\mu) \ : \ i = 1, \ldots, \dim P - \dim N\}$$

to be the set of possible $\lambda^{\perp}(x, v)$ attained for $v \in P$ and x is typical with respect to the measure μ. We also define

$$\Lambda_{\max}^{N,P} = \sup_{\mu \in \mathrm{Erg}(A)} \lambda_1^{N,P}(\mu)$$

and

$$\Lambda_{nat}^{N,P} = \lambda_1^{N,P}(\mu_{nat}).$$

In the case of symmetries we often obtain attractor within an invariant subspace N that is nontrivially contained within several distinct invariant subspaces P_1, P_2 etc. The L.E.s in different directions P_i can however in certain circumstances be related. In what follows T is some isotropy subgroup of Γ:

Theorem 4.1. *Suppose that $N = \mathrm{Fix}(T)$ and $N \subset P_k = \mathrm{Fix}(\Sigma_k)$, $k = 1, 2$ where $\Sigma_k \subset T$ are isotropy subgroups that are conjugate within T. Suppose that $A \subset N$ is an attractor. Then λ_i^{N,P_k}, $\Lambda_{\max}^{N,P_k}$ and Λ_{nat}^{N,P_k} are independent of k.*

Proof. We write $\Sigma = \Sigma_1$ so that $P_1 = \mathrm{Fix}(\Sigma)$. There is a $g \in T$ such that $P_2 = \mathrm{Fix}(g^{-1}\Sigma g) = g^{-1}\mathrm{Fix}(\Sigma) = g^{-1}P_1$. Now $gx = x$ and G has orthogonal action, so $|gy| = |y|$ for any y. Equivariance of f implies equivariance of the derivative $(Df(gx)gv = gDf(x)v)$. Hence for any $x \in N$ and $v \in P_1$ we have

$$\lambda(g^{-1}x, g^{-1}v) = \lim_{n\to\infty} \frac{1}{n} \log \frac{|Df^n(gx)gv|}{|gv|}$$

$$= \lim_{n\to\infty} \frac{1}{n} \log \frac{|gDf^n(x)v|}{|gv|} = \lambda(x, v).$$

This means that any L.E. λ_i^{N,P_1} is also a L.E. λ_i^{N,P_2} and vice versa. □

The previous result requires that the conjugating element g is in T. More generally one can require that g maps $\mathrm{Fix}(T)$ to itself. This implies that $g \in \mathrm{Norm}(T)$ where $\mathrm{Norm}(T) = \{h \in \Gamma \ : \ hT = Th\}$ is the normalizer of T. In this case the result above can be adapted as long as the measure μ_{nat} has symmetry g (a *symmetry on average* [19]). More precisely,

Theorem 4.2. *Suppose that $N = \mathrm{Fix}(T)$ and $N \subset P_k = \mathrm{Fix}(\Sigma_k)$ for $k = 1, \ldots, l$ where $\Sigma_k \subset T$ are related by $\Sigma_k = g_k^{-1}\Sigma g_k$ for some $g_k \in \mathrm{Norm}(T)$. Suppose that $A \subset N$ is an attractor with natural measure μ_{nat} invariant under action of all g_k. Then λ_i^{N,P_k}, $\Lambda_{\max}^{N,P_k}$ and Λ_{nat}^{N,P_k} are independent of k.*

Proof. This follows as for the previous result on noting that there is a g_k-invariant set of x with full μ_{nat}-measure that has the same L.E.s. at each point. □

4.1 Directions of Riddling

Recall that a basin of attraction $\mathcal{B}(A)$ is *riddled* (in M) [1] if for any $x \in A$ and $\delta > 0$ we have

$$\ell_M(\mathcal{B}(A) \cap B_\delta(x))\ell_M(\mathcal{B}(A)^c \cap B_\delta(x)) > 0$$

where $\ell_M(\cdot)$ denotes Lebesgue measure on M. Similarly, we say the basin of attraction of $A \subset N_0$ is *riddled in the direction* N_i for any invariant subspace N_i where $N_0 \subset N_i$ if

$$\ell_{N_i}(\mathcal{B}(A) \cap B_\delta(x))\ell_{N_i}(\mathcal{B}(A)^c \cap B_\delta(x)) > 0.$$

Observe that it is necessary for $\mathcal{B}(A) \cap N_i$ to have positive ℓ_{N_i}-measure in order to get riddling.

Suppose that A is a Milnor attractor such that $A \subset N_0 \subset N_1 \subset M$. If A is riddled in M then it need not be riddled in either N_i. In fact examples discussed in [1, 9] have attractors that are riddled in M but asymptotically stable and therefore unriddled in the largest linear subspace N that contains the attractor. What we emphasise here is that it may be unriddled in a *larger* invariant subspace. The same holds even if A is not an attractor but a chaotic saddle in M.

Suppose a Milnor attractor A in M has $\Lambda_{nat} < 0 < \Lambda_{\max}$ for the system restricted to an invariant subspace N' with $N \subset N' \subset M$. Then the same inequality holds for the full system and will imply riddling in the full basin, as long as A is an attractor in M.

One can find systems $f : M \to M$ with an invariant set A contained in an invariant subspace N such that A is an attractor in M but not in N. For example, consider the flow induced by the vector field

$$(\dot{x}, \dot{y}) = (x^3 - y^2x, x^2y - y^3) \tag{11}$$

shown in Fig. 5. This has an equilibrium at $(0, 0)$ that has the open basin of attraction given by $y^2 > x^2$. However, all points with $y^2 < x^2$ are repelled away to infinity. Thus the origin is an attractor in $\mathbb{R}^2$ but has trivial basin of attraction in the invariant subspace given by the x-axis. In this example this is a degeneracy caused by non-hyperbolicity of the fixed point at the origin. We expect that this behaviour cannot occur in sufficiently hyperbolic systems.

In applications where noise is highly directional, riddled basin attractors may have a degree of sensitivity of an attractor to anisotropic noise that is dependent on whether the noise is in directions in which the basin is unriddled or not as we will see later.

4.2 Internal and Other Riddling Transitions

Given any attractor A with a riddled basin in a system with several invariant subspaces, we can characterize this basin by examining the dimension $\tilde{d}$ of the

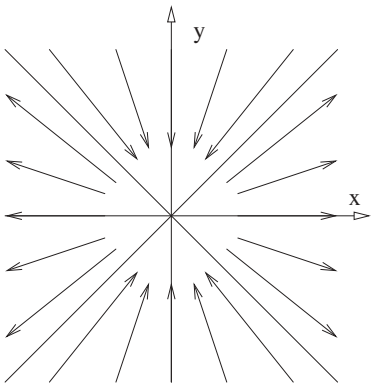


Fig. 5. The dynamics near the non-hyperbolic fixed point at the origin $(0,0)$ for (11) such that it is a Milnor attractor for the full system but not for the system restricted to the invariant subspace $(x,0)$

largest invariant subspace N containing A such that the basin of $\mathcal{B}(A) \cap N$ is unriddled. Note that $\dim(A) \leq \tilde{d} \leq \dim(M)$. Any change in $\tilde{d}$ we term an *internal riddling transition*. It is clear that such transitions will occur in higher-dimensional systems; as A loses asymptotic stability in more directions the index $\tilde{d}$ will decrease.

Similarly one can apply the standard riddling bifurcation criteria in each subspace to predict parameter values when internal riddling transitions will occur; see [16, 48, 51, 63, 81]. Such transitions will typically be rather unclear on varying system parameters if the chaotic attractors are not structurally stable; only by examining normal parameters of the system such that the dynamics on the attractor is left unchanged can one hope to find internal riddling transitions appearing as codimension one transitions. Similar transitions where riddling bifurcations are replaced by blowout bifurcations will also occur in such systems; see for example the systems studied in [5].

4.3 A Numerical Example

If we examine the map (8) for parameter values that are intermediate between strong and weak coupling one may find a wide variety of attractors of different symmetries that are multi-stable. One can also find, for example, (i) chaotic saddles that have basins that are riddled within certain invariant subspaces (ii) attractors that are riddled in some directions but not others and hence (iii) anisotropic bubbling response to anisotropic noise.

Here we consider an attractor in

$$\mathrm{Fix}(S_2 \times S_2) = \{(p,p,q,q) \ : \ p,q \in \mathbb{R}\}.$$

that occurs when $a = 1.71$ and $\epsilon = 0.15$. Figure 6(i) shows a time series on this attractor for an initial condition very close to

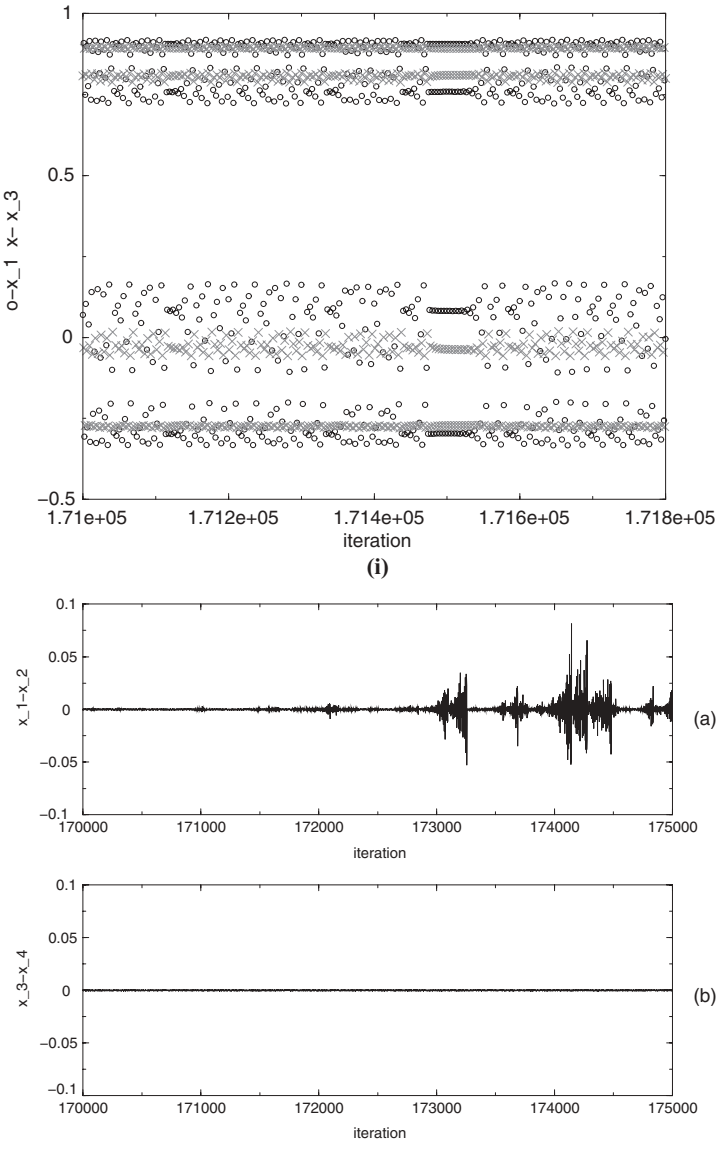


Fig. 6. Anisotropic bubbling behaviour caused by anisotropic riddling of the basin of an attractor for a system of four globally coupled maps (8) that lies in an invariant subspace with symmetry $S_2 \times S_2$. (i) shows time-series for the noise-free system; the circles show the values of x_1 and x_2 while the crosses show the values of x_3 and x_4. In (ii) the attractor in (i) is subject to very low amplitude noise. In (iia) the noise is added in only the x_1 direction giving rise to bubbling in this direction. In (iib) it is added in only the x_3 direction giving a stable response, due to the attractor not being riddled in this direction

$$(p, p, q, q) = (0.00292, 0.00292, 0.8004, 0.8004)$$

and no added noise; $\sigma_j = 0$ for all j. On addition of noise (iia) shows $x_1 - x_2$ for the same initial condition but with $\sigma_1 = 10^{-5}$. The large deviations away from $x_1 = x_2$ indicative of bubbling in the direction $(1, 0, 0, 0)$ and hence indicate that the basin of attraction of this attractor is riddled to perturbations into the invariant subspace (p, q, r, r). By contrast, (iib) shows $x_3 - x_4$ for the same initial condition but with $\sigma_3 = 10^{-5}$. In this case there is apparently stable response indicating that the basin is not riddled into the invariant subspace (p, p, q, r). Note that (i) shows that the statistics of x_1 and x_3 are quite different; the natural invariant measure associated with this attractor is not invariant under the transformation

$$(x_1, x_2, x_3, x_4) \mapsto (x_3, x_4, x_1, x_2).$$

If this symmetry did leave the attractor invariant then by applying Theorem 4.2 either both or neither of the directions $(1, 0, 0, 0)$ and $(0, 0, 1, 0)$ would be riddled. Numerical calculations indicate that the transverse L.E. in the direction $(1, 0, 0, 0)$ is approximately -7.8×10^{-4} whereas it is approximately -0.2842 in the direction $(0, 0, 1, 0)$; this agrees with the anisotropic bubbling observations.

5 Some Open Problems

In the final section we highlight some open problems and themes of interest related to riddled basins.

5.1 Unfolding of Blowout and Riddling

One of the most intriguing aspects of blowout bifurcation and associated transitions to riddled basins are the observation, noted first by [64] in many examples one can clearly classify the nonlinear stability of the system into one of two scenarios: either subcritical, where a riddled basin attractor loses its basin at blowout to become a repellor, or supercritical, where an attractor is locally but not globally riddled before blowout, and after blowout there is on-off intermittency [37, 69, 70] or a stuck-on attractor [3]. (See [67, 68, 84] for more early work on loss of chaotic synchronization)

By analogy with bifurcation of fixed points, it would be nice to understand precisely what determines criticality at blowout and hence whether basins of attraction are locally or globally riddled. The map (1,2,3) was studied in [6] where it was noted that for this map, the two scenarios could be distinguished subcritical (resp. supercritical) if the *essential basin of attraction* of A was zero (resp. positive) measure at the point of blowout.[4] In the general case this

[4] The essential basin of A is defined as the set of points y whose trajectories visit any neighbourhood of A with positive frequency; i.e. such that $\limsup_{n\to\infty} \frac{1}{n}\#\{0 \le k < n \ : \ f^k(y) \in U\} > 0$ for all U neighbourhoods of A.

amounts to a conjecture. Other work has examined *absorbing areas* [21, 52] and weak attractors [16] to give criteria for supercriticality; for example in [16] the presence of a larger weak attractor containing a riddled basin attractor seems to indicate that blowout will be supercritical.

Generally, can one find conditions that ensure that a "scenario" occurs at blowout and if we can do this, find conditions so that the classification of the blowout into a scenario is computable in some sense? So far, it seems that is may only be possible to answer this in a very limited sense; if so, why can one nevertheless observe this scenario in many simulations?

5.2 Riddling for Infinite Dimensions

Many mathematical models that arise include temporal delays and/or spatial extension in such a way that the phase space on which the dynamics occurs is unavoidably infinite dimensional. So far there is no corresponding concept for riddled basin, or indeed Milnor attractor, that works in cases where a nice background measure such as Lebesgue is not available. It should be possible to use ideas such as prevalence [39] in such cases to make some progress but this has not yet been done to my knowledge.

5.3 Pseudo-Riddled and δ-Riddled Basins

A strange feature of riddling is that in cases that one can prove it exists it is often hard to find, while in cases where can prove it does not exist, numerics often seem to say the opposite. For example, in [49] examples of pseudo-riddled basins are presented. In [12] the concept of δ-riddling was used to give a numerical profile of pseudo-riddling by saying that a set is δ-riddled if (6) holds for a given δ; the set is then riddled in the usual sense if it is δ-riddled for all $\delta > 0$. It would be nice to gain a better theoretical understanding of these effects and their consequences, though any simple minded approach clearly has the disadvantage of not being invariant under coordinate changes.

5.4 Genericity of Riddled Basins

In dissipative dynamics that has no imposed invariant subspaces, it is the belief of the writer (and others) that riddled basin attractors do not arise except in exceptional circumstances, namely when there are invariant subspaces forced by system symmetries or for example by other constraints [72]. Can one prove any meaningful results in this direction? An important part of this questions is to understand the prevalence of the appearance of chaotic attractors; see for example [20, 22, 40, 50, 57].

The appearance of a minimal Milnor attractor that has a partially riddled basin was suggested in [16] to be a degenerate case; can one prove a sense in which this is for example non-generic in a class of smooth systems?

5.5 Unstable Attractors

There has been some very interesting recent work looking at dynamical systems of globally delay pulse coupled oscillator systems, motivated by simple models of neural systems. These systems can possess attractors that are an extreme case of riddling; there is a neighbourhood of the attractor that has zero measure intersection with the attractor's basin. Such *unstable attractors* have been found to be quite widespread in the dynamics of certain systems [80]. It is a challenge to understand constraints on the appearance of such attractors and their characteristic properties.

5.6 Non-Ergodic Attractors and Riddling

In most standard examples where one can prove the existence of riddled basins the attractor in question is ergodic, namely it possesses an SRB measure whose support is the attractor. Nonetheless, in the presence of symmetries it has been recognised for some time that for many cases one can find structurally stable attractors that are not ergodic; in particular robust attracting heteroclinic cycles [35, 46, 47] between saddle equilibria, periodic orbits or even chaotic saddles [4, 27, 31]. These attractors are non-ergodic and not even transitive (no dense orbits); they consist of chains of connecting orbits and invariant sets [13, 14].

As a particular example of this in a finite coupled map lattice is the system with one-directional coupling considered in [15]:

$$X_{n+1}^k = f(X_n^k)\, e^{-\gamma X_n^{k-1}}$$

where $n \in \mathbb{N}$, $k \in \mathbb{Z}$ is taken modulo 3 and $f(x) = rx(1-x)$. Due to the presence of invariant subspaces $X^k = 0$ (for any k) if we choose $2 < r < 4$ and γ large enough this system can be shown to have open sets of spatially periodic initial conditions that are attracted to states with no convergence of ergodic averages; see [15] for more details.

For a skew product system of a chaotic attractor forcing a robust heteroclinic cycle one can exhibit examples where the cycle has a locally riddled basin even though the attractor is non-ergodic [4]. How general is this?

5.7 Other Problems Related to Riddled Basins

Some researchers have implied that there are connections between the computability of riddled basins and decidability [65, 76]. It would be of interest to know if riddling may help one to understand the nature of chaotic behaviour in area-preserving maps such as the standard map, or in other area-preserving maps; for example Fig. 7 shows in black the invariant set of points that approach arbitrarily close the discontinuity for an area-preserving map that arises in signal processing; see [10, 12] for a discussion of this and generalization of the results such as Theorem 2.1 to area-preserving discontinuous

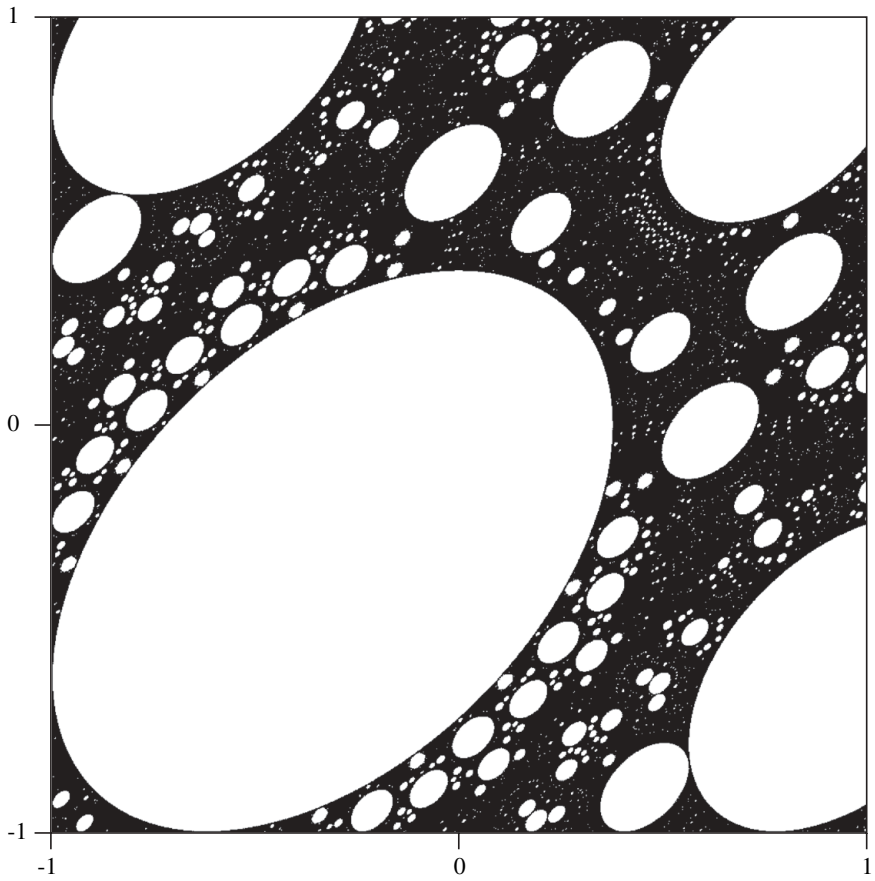


Fig. 7. The *black* region shows the set of points within the square $[-1, -7/8]^2$ whose orbits accumulate on the discontinuity for the area-preserving map $(x, y) \mapsto (y, g(-x + 0.9y))$ where $g(x) = x$ for $x \in [-1, 1)$ and $g(x + 2) = g(x)$. It is an open question whether this set is riddled; the white regions are packed with invariant curves centred on periodic points and the map can be viewed as a planar piecewise isometry

maps. It would be of great interest to be able to determine if the set shown in this figure is riddled or not. Other problems include getting a better understanding of properties of intermingled basins.

Acknowledgements

I would like to thank many collaborators over the years who have helped and stimulated my understanding of riddling effects. In particular I would like to thank P. Aston, M. Breakspear, J. Buescu, E. Covas, M. Field,

X. Fu, I. Melbourne, M. Nicol, R. Roy, I. Stewart, R. Tavakol and J. Terry. I would also like to thank J.-R. Chazottes and B. Fernandez for organizing the school-forum CML2004 that provided the occasion to prepare these notes.

References

1. J.C. Alexander, J.A. Yorke, Z. You and I. Kan,: Riddled Basins. *Int. Journal of Bifurcations and Chaos* **2** (1992), 795–813. 193, 197
2. J.C. Alexander, B.R. Hunt, I. Kan, and J.A. Yorke, Intermingled basins for the triangle map Ergodic Theory Dynam. Systems **16** (1996), no. 4, 651–662. 193
3. P. Ashwin: Attractors stuck on to invariant subspaces *Phys. Lett. A* **209**, 338–344 (1995). 200
4. P. Ashwin: Cycles homoclinic to chaotic sets; robustness and resonance *Chaos* **7**, 207–220 (1997). 202
5. P. Ashwin and P.J. Aston: Blowout bifurcations of codimension two. *Physics Letters A* **244**, 261–270 (1998). 198
6. P. Ashwin, P.J. Aston, P. and M. Nicol: On the unfolding of a blowout bifurcation. *Physica D* **111**, 81–95 (1998). 183, 184, 185, 200
7. P. Ashwin and M. Breakspear: Anisotropic properties of riddled basins. *Phys. Lett. A* **280**, 139–145 (2001). 181, 193, 195
8. P. Ashwin, J. Buescu and I.N. Stewart: Bubbling of attractors and synchronisation of oscillators. *Phys. Lett. A* **193**, 126–139 (1994). 189, 193
9. P. Ashwin, J. Buescu and I. Stewart: From attractor to chaotic saddle: a tale of transverse instability. *Nonlinearity* **9**, 703–737 (1996). 189, 192, 193, 195, 197
10. P. Ashwin, W. Chambers and G. Petkov: Lossless digital filter overflow oscillations; approximation of invariant fractals. *Int. J. Bifur. Chaos* **7**, 2603–2610 (1997). 202
11. P. Ashwin, E. Covas and R.K. Tavakol: Transverse instability for non-normal parameters. *Nonlinearity* **12**, 563–577 (1999). 193, 195
12. P. Ashwin, X.-C. Fu and J.R. Terry: Riddling and invariance for discontinuous maps preserving Lebesgue measure. *Nonlinearity* **15**, 633-645 (2002). 188, 201, 202
13. P. Ashwin and A.M. Rucklidge: Cycling chaos: its creation, persistence and loss of stability in a model of nonlinear magnetoconvection. *Physica D* **122**, 134–154 (1998). 191, 202
14. P. Ashwin, A.M. Rucklidge and R. Sturman: Cycling attractors of coupled cell systems and dynamics with symmetry. In: *Proceedings of Workshop on Synchronization, Crimea, 2002.* Eds A. Pikovsky and Y. Maistrenko, Kluwer (2003). 191, 202
15. P. Ashwin and R. Sturman: Decelerating defects and non-ergodic critical behaviour in a unidirectionally coupled map lattice. *Phys Letters A* **309**, 423–434 (2003). 202
16. P. Ashwin and J.R. Terry: On riddling and weak attractors. *Physica D* **142**, 87–100 (2000). 183, 187, 188, 189, 193, 198, 201
17. P.J. Aston and M. Dellnitz: Symmetry breaking bifurcation of chaotic attractors. *Int. J. Bif. Chaos* **5**, 1643-1676 (1995). 191
18. V. Baladi and M. Viana: Strong stochastic stability and rate of mixing for unimodal maps *Ann. Sci. École Norm. Sup. (4)* **29**, 483–517 (1996). 190

19. E. Barany, M. Dellnitz and M. Golubitsky: Detecting the symmetry of attractors. *Physica D* **67**, 66–87 (1993). 191, 196
20. E. Barreto, B.R. Hunt, C. Grebogi and J. Yorke: From high dimensional chaos to stable periodic orbits: the structure of parameter space. *Phys. Rev. Lett.* **78**, 4561–4564 (1997). 195, 201
21. G.-I. Bischi and L. Gardini: Role of invariant and minimal absorbing areas in chaos synchronization. *Phys. Rev. E* **58**, 5710–5719 (1998). 201
22. M. Benedicks and L. Carleson: On iterations of $1 - ax^2$ on $[-1, 1]$. *Ann. Math.* (2) **122**, 1–25 (1985). 201
23. J. Buescu: *Exotic Attractors.* (Birkhäuser Verlag, Basel 1997). 192
24. Y. Cao: A note about Milnor attractor and riddled basin. *Chaos, Solitons and Fractals* **19**, 759–764 (2004). 183, 193
25. P. Chossat and M. Golubitsky: Symmetry increasing bifurcation of chaotic attractors, *Physica D* **32**, 423–436 (1988). 191
26. E. Covas, P. Ashwin and R.K. Tavakol: Non-normal parameter blowout bifurcation in a truncated mean field dynamo. *Phys. Rev. E* **56**, 6451–6458 (1997). 195
27. M. Dellnitz, M. Field, M. Golubitsky, A. Hohmann and J. Ma: Cycling chaos, *IEEE Trans. Circuits and Systems-I.* **42**, 821–823 (1995). 191, 202
28. M. Ding and W. Yang: Observation of intermingled basins in coupled oscillators exhibiting synchronized chaos. *Phys. Rev. E* **54**, 2489–2494 (1996). 193, 194
29. J.-P. Eckmann and D. Ruelle: Ergodic Theory of Chaos and Strange Attractors. *Rev. Mod. Phys.* **57**, 617–656 (1985). 182, 192
30. J.D. Farmer: Sensitive dependence on parameters in nonlinear dynamics. *Phys. Rev. Lett.* **55**, 351–354 (1985). 187
31. M. Field. *Lectures on Dynamics, Bifurcation and symmetry.* Pitman Research Notes in Mathematics, **356**, Pitman (1996). 191, 202
32. K. Falconer. *Fractal Geometry* (John Wiley & sons, Chichester 1990). 187, 188
33. M. Golubitsky and I. Stewart. *The symmetry perspective* (Birkhäuser Verlag, Boston 2002). 191
34. C. Grebogi, S.W McDonald, E. Ott & J.A. Yorke. Exterior dimensions of fat fractals *Phys. Lett. A* **110**:1–4 (1985). 187, 188
35. J. Guckenheimer and P. Holmes. Structurally stable heteroclinic cycles, *Math. Proc. Camb. Phil. Soc.* **103**:189–192 (1988). 202
36. M. Hasler, Y. Maistrenko and O. Popovych: Simple example of partial synchronization of chaotic systems. *Phys. Rev. E* **58**, 6843–6346 (1998). 193
37. J.F. Heagy, N. Platt and S.M. Hammel: Characterization of on-off intermittency. *Phys. Rev. E* **49**, 1140–1150 (1994). 200
38. J.F. Heagy, T. Carroll and L. Pecora: Experimental and numerical evidence for riddled basins in coupled chaotic systems. *Phys. Rev. Lett.* **73**, 3528–31 (1994). 193
39. B.R. Hunt: The Prevalence of Continuous Nowhere Differentiable Functions *Proc. Amer. Math. Soc.* **122**, 711–717 (1994). 201
40. M. Jakobson: Absolutely continuous invariant measures for one parameter families of one-dimensional maps, *Comm. Math. Phys.* **81**, 39–88 (1981). 201
41. I. Kan: Open sets of diffeomorphisms having two attractors, each with an everywhere dense basin. *Bull. Am. Math. Soc.* **31**, 68–74 (1994). 193
42. K. Kaneko: On the strength of attractors in a high-dimensional system: Milnor attractor network, robust global attraction, and noise-induced selection. *Physica D* **124**, 322–344 (1998). 191
43. K. Kaneko: Mean field fluctuation of a network of chaotic elements. *Physica D* **55**, 368–384 (1992).

44. J.L. Kaplan and J.A. Yorke: Chaotic behaviour of multidimensional difference equations. In *Functional differential equations and approximations of fixed points*, (H.-O. Peitgen and H.-O. Walther, eds) Springer Lecture Notes in Maths **730** (1979). 189
45. L. Kocarev, A. Shang and L.O. Chua: Transitions in dynamical regimes by driving: a unified method of control and synchronisation of chaos. *Intl. J. Bifurcation and Chaos* **3**, 479–483 (1993).
46. M. Krupa and I. Melbourne: Asymptotic stability of heteroclinic cycles in systems with symmetry, *Erg. Th. Dyn. Sys.* **15**, 121–147 (1995). 202
47. M. Krupa: Robust heteroclinic cycles, *Journal of Nonlinear Science* **7**, 129–176 (1997). 202
48. Y.-C. Lai, C. Grebogi, J.A. Yorke and S.C. Venkataramani: Riddling bifurcation in chaotic dynamical systems. *Phys. Rev. Lett.* **77**, 55–58 (1996). 198
49. Y.-C. Lai: Pseudo-riddling in chaotic systems. *Phys. D* **150**, 1–13 (2001). 201
50. A. Lasota and J.A. Yorke: On the existence of invariant measures for piecewise monotonic transformations. *Trans. Am. Math. Soc.* **186**, 481–488 (1973). 201
51. Yu.L. Maistrenko, V.L. Maistrenko, A. Popovich and E. Mosekilde: Transverse instability and riddled basins in a system of two coupled logistic maps. *Phys. Rev. E* **57**, 2713–2724 (1998). 198
52. Yu.L. Maistrenko, V.L. Maistrenko, A. Popovich and E. Mosekilde: Role of the absorbing area in chaotic synchronization. *Phys. Rev. Lett.* **80**, 1638–1641 (1998). 201
53. R. Mañé: *Ergodic Theory and Differentiable Dynamics.* (Springer-Verlag, Berlin 1987). 192
54. I. Melbourne: An example of a non-asymptotically stable attractor. *Nonlinearity* **4**, 835–844 (1991). 187
55. I. Melbourne, M. Dellnitz and M. Golubitsky: The Structure of Symmetric Attractors, *Arch. Rational. Mech. Anal.* **123**, 75–98 (1993). 191
56. J. Milnor: On the concept of attractor. *Commun. Math. Phys.* **99**, 177–195 (1985); Comments *Commun. Math. Phys.* **102**, 517–519 (1985). 182, 183
57. L. Mora and M. Viana: Abundance of strange attractors. *Acta Math.* **171**, 1–71 (1993). 201
58. E. Mosekilde, Y. Maistrenko and D. Postnov: *Chaotic synchronization. Applications to living systems.* (World Scientific Publishing, River Edge NJ, 2002). 181
59. H. Nakajima and Y. Ueda: Riddled basins of the optimal states in learning dynamical systems. *Physica D* **99**, 35–44 (1996). 193
60. V.I. Oseledec: A multiplicative ergodic theorem: Liapunov characteristic numbers for dynamical systems. *Trans. Mosc. Math. Soc.* **19**, 197–231, (1968). 192
61. E. Ott: *Chaos in Dynamical Systems* (Second edition) (Cambridge University Press 2002). 182
62. E. Ott, J.C. Sommerer, J. Alexander, I. Kan and J.A. Yorke: Scaling behaviour of chaotic systems with riddled basins. *Phys. Rev. Lett.* **71**, 4134–4137 (1993). 188, 189
63. E. Ott, J.C. Sommerer, J. Alexander, I. Kan and J.A. Yorke: A transition to chaotic attractors with riddled basins. *Physica D* **76**, 384–410 (1994). 198
64. E. Ott and J.C. Sommerer: Blowout bifurcations: the occurrence of riddled basins and on-off intermittency. *Phys. Lett. A* **188**, 39–47 (1994). 200
65. M.W. Parker: Undecidability in $\mathbb{R}^n$; riddled basins, the KAM tori and the stability of the solar system. *Philos. Sci.* **70**, 359–382 (2003). 202
66. Y.B. Pesin: Characteristic Liapunov exponents and smooth ergodic theory. *Russian Math. Surv.* **32**, 55–114 (1978). 192

67. A.S. Pikovsky: On the interaction of strange attractors. *Z. Phys. B*, **55**, 149–154 (1984). 200
68. A.S. Pikovsky and P. Grassberger: Symmetry breaking of coupled chaotic attractors. *J. Phys. A*, **24**, 4587–4597 (1991). 200
69. N. Platt, E. Spiegel and C. Tresser: On-off intermittency; a mechanism for bursting. *Phys. Rev. Lett.* **70**, 279–282 (1993). 200
70. N. Platt, S.M. Hammel and J.F. Heagy: Effects of additive noise on on-off intermittency. *Phys. Rev. Lett.* **72**, 3498–3501 (1994). 200
71. C. Pugh and M. Shub: Ergodic attractors. *Trans. Am. Math. Soc.* **312**, 1–54 (1989). 182, 192
72. D.A. Rand, H. Wilson and J. McGlade: Dynamics and evolution: evolutionarily stable attractors, invasion exponents and phenotype dynamics. *Phil. Trans. R. Soc. Lond. B*, **343**, 261–283 (1994). 201
73. D. Ruelle: Analyticity properties of the characteristic exponents of random matrix products, *Advances in Math.* **32**, 68–80 (1979). 195
74. D. Ruelle: *Chaotic evolution and strange attractors*, (Lezione Lincee, Cambridge University Press, 1989). 195
75. R.J. Sacker and G.R. Sell: The spectrum of an invariant submanifold. *J. Diff. Equations* **38**, 135–160 (1980). 192
76. A. Saito and K. Kaneko: Inaccessibility and indecidability in computation, geometry and dynamical systems. *Physica D* **155**, 1–33 (2001). 202
77. S.H. Strogatz: *Synch: the emerging science of spontaneous order*, (Hyperion Press 2003). 181
78. K. Sygmund: Generic properties of invariant measures for Axiom A diffeomorphisms. *Inv. Math.* **11**, 99–109 (1970). 193
79. A.V. Taborev, Y.L. Maistrenko, and E. Mosekilde: Partial synchronization in a system of coupled logistic maps. *Int. J. Bifurcation and Chaos* **10**, 1051–1066 (2000). 191
80. M. Timme, F. Wolf, and T. Geisel: Prevalence of unstable attractors in networks of pulse-coupled oscillators, *Phys. Rev. Lett.* **89**, 154105 (2002). 202
81. S.C. Venkataramani, B.R. Hunt, E. Ott, D.J. Gauthier and J.C. Bienfang: Transitions to bubbling of chaotic systems. *Phys. Rev. Lett.* **77**, 5361–5364 (1996). 198
82. R. L. Wheeden and A. Zygmund: Measure and integral. (Marcel Dekker Inc, New York 1977). 188
83. M. Woltering and M. Markus: Riddled basins of coupled elastic arches. *Phys. Lett. A* **260**, 453–461 (1999). 193
84. H. Yamada and T. Fujisaka: Stability theory of synchronised motion in coupled-oscillator systems. *Prog. Theor. Phys.* **70**, 1240–1248 (1984). 200

Some Topological Properties of Lattice Dynamical Systems

V. Afraimovich

Universidad Autónoma de San Luis Potosí, Alvaro Obregón 64, San Luis Potosí, SLP 78000, México
valentin@cactus.iico.uaslp.mx

1 Introduction

Phenomenological models of motions in media with dissipation in the form of lattices of coupled ordinary differential equations or maps appeared about 50 years ago and since then play an important role in study of dynamical properties of systems in material science, fluid dynamics, chemistry, image processing, biology, etc [1]. We will call them Lattice Dynamical Systems (LDS), see below. A special class of LDS, the so called Coupled Map Lattices (CML), has been introduced almost simultaneously in [2, 3, 4, 5] and, mainly because of convenience of numerical simulations, became a basic model in the field [6]. Beginning with [7], many rigorous mathematical results have been obtained concerning topological and ergodic features of LDS and CML.

In this chapter we try to describe some rigorous results and ideas related to topological properties (see also [8]).

2 LDS with Discrete Time. Definitions

Being models of extended media LDS have to be, in principle, infinite-dimensional dynamical systems. At the same time, there are many models where one has only a finite number of coupled active elements or even just a few. In this chapter we will focus on the infinite dimensional models; some problems of correspondence between finite and infinite-dimensional cases will be discussed at the end of the chapter.

2.1 The Phase Space

The phase space of LDS is the set of all configurations

$$\mathcal{B} = \{u \,|\, u = \{u_j\},\ u_j \in \mathbb{R}^p,\ j \in \mathbb{Z}^d,\ \| u \| < \infty\}$$

V. Afraimovich: *Some Topological Properties of Lattice Dynamical Systems*, Lect. Notes Phys. **671**, 153–179 (2005)
www.springerlink.com

where $d \geq 1$ is the dimension of the lattice, $p \geq 1$ is the dimension of the range space of the function of state u_j and the norm $\|\cdot\|$ is usually taken to be one of the following ones ($|u_j|$ is the length of the vector u_j)

- the l^∞-norm, $\quad \| u \|_\infty = \sup_{j\in\mathbb{Z}^d} |u_j|$.
- the l^2 norm, $\quad \| u \|_2 = (\sum_{j\in\mathbb{Z}^d} |u_j|^2)^{\frac{1}{2}}$.
- the l^2_q norm, $\quad \| u \|_q = (\sum_{j\in\mathbb{Z}^d} \frac{|u_j|^2}{q^{|j|}})^{\frac{1}{2}}$ where $q > 0$ is a fixed number.

We note that $l^2_1 = l_2$. It is simple to see that $\mathcal{B}$ is a Hilbert space in the l^2_q-norm and a Banach space in the l^∞-norm. The l^∞-norm is consistent with the uniform topology while the l^2_q-norm, $q > 1$, is consistent with the product topology. Both are important because they reflect different features of lattice dynamics in the infinite dimensional space. Therefore, we use both these norms through all the chapter. Of course, sometimes people use other norms [24].

2.2 The Evolution Operator

In the definition below it is assumed that the local dynamics at the site j depends only on finitely many neighboring sites. Of course, there are models where coupling depends on infinitely many states, but in this chapter we consider the simplest situation.

The evolution operator $\mathcal{F} : \mathcal{B} \to \{\mathbb{R}^p\}^{\mathbb{Z}^d}$ generating LDS $u' = \mathcal{F}(u)$ is defined by

$$u'_j = F\left(\{u_j\}^s\right) \tag{1}$$

where $\{u_j\}^s = \{u_i \,|\, |i-j| \leq s\}$, $|j| = \max\{|j_1|, \ldots, |j_d|\}$ or $|j| = |j_1| + \cdots |j_d|$ and $F : (\mathbb{R}^p)^{(2s+1)^d} \to \mathbb{R}^p$ is a smooth function. It follows that $\mathcal{F}(\mathcal{B}) \subset \mathcal{B}$ (in the case of l^2_q-norm, $0 < q \leq 1$, it is always assumed that $F(\{0\}^s) = 0$). To avoid non-essential technicalities let us assume from the very beginning that F is C^r, $r \geq 2$, and there exists a constant $M > 0$ such that

$$\left|\frac{\partial F}{\partial u_i}\right| \leq M, \qquad \left|\frac{\partial^2 F}{\partial u_i \partial u_j}\right| \leq M \tag{2}$$

for any collection $\{u_0\}^s \in (\mathbb{R}^p)^{(2s+1)^d}$. Let us remark that these assumptions are not as restrictive as they seem. In fact people are only interested in bounded configurations but not in their asymptotic behavior as $|u_j|$ goes to infinity.

The following statements hold:

- In the l^∞-norm and in the l^2_q-norm, $0 < q \leq 1$, $\mathcal{F}$ is C^r.
- In the l^q-norm, $q > 1$, $\mathcal{F}$ is Lipschitz-continuous with the constant $L = M(2s+1)^{\frac{3}{2}d} q^{\frac{s}{2}}$ [9].

Definition 2.1. *The dynamical system* $(\mathcal{F}, \mathcal{B})$ *will be referred to as a* lattice dynamical system (LDS) *and we will write*

$$u(n+1) = \mathcal{F}(u(n)) \;\text{ or }\; u_j(n+1) = F(\{u_j(n)\}^s)\ . \tag{3}$$

2.3 Translational DS

The group of translations $\{S_{j_0}\}_{j_0 \in \mathbb{Z}^d}$ acts in the space $\mathcal{B}$ as follows:

$$(S_{j_0} u)_j = u_{j+j_0} \text{ if } u = \{u_j\}\ .$$

Each element S_{j_0} is a linear bounded operator (an isometry in the l^2-norm or in the l^∞-norm). Clearly each map S_{j_0} commutes with the evolution operator $\mathcal{F}$. Sometimes the dynamical system $\left(\{S_{j_0}\}_{j_0 \in \mathbb{Z}^d}, \mathcal{B}\right)$ is called the *translational dynamical system (TDS).*

2.4 Examples

Among the most important examples are the following ones

- Discrete version of the reaction-diffusion equation:
  - One dimensional case

$$u_j(n+1) = u_j(n) + \alpha f(u_j(n)) + \epsilon\left(u_{j-1}(n) - 2u_j(n) + u_{j+1}(n)\right)\ . \tag{4}$$

  - Two dimensional case

$$\begin{aligned} u_{j,k}(n+1) = u_{j,k}(n) + \alpha f(u_{j,k}(n)) + \epsilon(u_{j-1,k}(n) \\ - 4u_{j,k}(n) + u_{j-1,k}(n) + u_{j,k-1}(n) + u_{j,k+1}(n)) \end{aligned} \tag{5}$$

    where $j, k \in \mathbb{Z}$, $n \in \mathbb{Z}_+$ and α, ϵ are non-negative parameters.
- Linearly coupled map

$$\mathcal{F}(u) = L \circ \mathcal{G}(u) \tag{6}$$

  where $(\mathcal{G}u)_s = f(u_s)$, $s \in \mathbb{Z}^d$, L is a linear bounded operator and f is a non-linear function. For example, the diffusively coupled maps are very popular in the literature:

$$u_j(n+1) = f(u_j(n)) + \epsilon\left(f(u_{j-1}(n)) - 2f(u_j(n)) + f(u_{j+1}(n))\right) \tag{7}$$

2.5 Simplest Orbits of LDS

The following types of orbits are the simplest ones and should be studied at first:

- Steady-state solutions or fixed points.
  They do not depend on time and satisfy the following equation:

$$u_j = F\left(\{u_j\}^s\right)\ . \tag{8}$$

- Spatially-homogeneous solutions.
  They do not depend on spatial coordinates and satisfy the equation:

$$v(n+1) = F\left(\{v(n)\}^s\right) \ . \tag{9}$$

- Traveling wave solutions with integer wave-numbers and velocities.
  These solutions depend on a linear combination of temporal and spatial coordinates and have the form:

$$u_j(n) = \psi\left(l \cdot j + mn\right) \tag{10}$$

  where $l \in \mathbb{Z}^d$ is an integer wave vector, $l \cdot j$ is the scalar product and $\psi : \mathbb{Z} \to \mathbb{R}^p$ is a vector function.
  In this chapter we will discuss mainly these configurations.

3 Fixed Points of LDS

One of the important problems related to fixed points is the problem of stability. We describe two approaches to solve it.

3.1 Finite Dimensional Description

In this subsection we consider the only case $d = 1$. Then the (8) can be rewritten as:

$$u_j = F(u_{j-s}, \ldots, u_{j+s}), \ \ j \in \mathbb{Z} \, . \tag{11}$$

Thoroughly we impose the following assumption

Assumption 3.1 $\frac{\partial F}{\partial u_{j-s}} \neq 0$, $\frac{\partial F}{\partial u_{j+s}} \neq 0$ *for any collection* $\{u_j\}^s \in \mathbb{R}^{2s+1}$.

It implies that (11) can be solved with respect to u_{j+s}, i.e., because of the implicit function theorem, 11 is equivalent to

$$u_{j+s} = g(u_{j-s}, \ldots, u_{j+s-1}), \ \ j \in \mathbb{Z} \, . \tag{12}$$

Introduce the notations

$$u_{-s} = x_1, \ u_{-s+1} = x_2, \ \ldots \ , \ u_{s-1} = x_{2s},$$

and the map $G : \mathbb{R}^{2s} \to \mathbb{R}^{2s}$ defined by

$$G(x) = x' \tag{13}$$

where $x = (x_1, \ldots, x_{2s})$, $x' = (x'_1, \ldots, x'_{2s})$ and

$$x'_1 = x_2 \ , \ldots \ , \ x'_{2s-1} = x_{2s}, \ x'_{2s} = g(x_1 \ldots, x_{2s}) \, .$$

It directly follows from the definition that every configuration $u = \{u_j\}$ satisfying (11) corresponds to the orbit $x^j = G^j(x^0)$ of the dynamical system generated by G. Indeed, define a correspondence χ as follows. Given $x^0 \in \mathbb{R}^{2s}$ let $\left(\chi(x^0)\right)_j = (x^j)_s$ be the s-coordinate of the vector x^j. Denote by $\Lambda \in \mathbb{R}^{2s}$ the set of all points of all bounded orbits of dynamical system (13). It can be easily shown that the following statement hold.

Lemma 3.1. *For any norm in $\mathcal{B}$, the map $\chi : \Lambda \to \mathcal{B}$ is injective and satisfies*

$$\chi \circ G = S_1 \circ \chi \tag{14}$$

where S_1 is a generator of translational dynamical system; in the l_q^2-norm, χ is a homeomorphism for any $q > 1$; and there exists $q_0 > 1$ such that for any $q \geq q_0$, $\Lambda = \mathbb{R}^{2s}$ and χ is a diffeomorphism.

The first statement follows from the definition of χ and the fact that, because of Assumption 3.1, G is invertible. Now, if x^0 and y^0 are sufficiently close, then x^j and y^j are close too for $|j| < N$, so that $\chi(x)$ and $\chi(y)$ are close in the l_q^2-norm, $q > 1$, which implies the continuity of χ. Differentiability can be proved by direct calculations.

Thus, to list fixed point of $\mathcal{F}$ one needs to describe all bounded orbits of dynamical system (13).

Example 3.1. Consider the system (4). Then (12) becomes

$$u_{j+1} = 2u_j - \frac{\alpha}{\epsilon} f(u_j) - u_{j-1} \tag{15}$$

and the dynamical system (13) reads

$$x_1' = x_2, \quad x_2' = 2x_2 - \frac{\alpha}{\epsilon} f(x_2) - x_1 \, . \tag{16}$$

It can be shown (see for instance [10]) that for $f(x) = bx(x^2 - 1)$ there is a G-invariant set Λ such that $G|_\Lambda$ is topologically conjugated to the Bernoulli shift with three symbols. It was also proved in [11] that the trajectories in Λ are the only bounded ones provided that b is large enough.

3.2 Stability: Finite Dimensional Approach

There are many results about stability of fixed points in LDS, see for instance [13, 14] and references therein. In this chapter we describe two methods.

Following the ideas of L. Glebsky [12] we present an approach to study stability in the l^∞-norm for $d = 1$. Let $u^* = \{u_j^*\}$ be a configuration satisfying (11). We are interested in the spectrum of the linearization operator $\mathcal{L}$ of $\mathcal{F}$ at u^*, defined by $\xi' = \mathcal{L}\xi$, explicitely given by

$$\xi_j' = A_j \left(\{\xi_j\}^s\right) \tag{17}$$

where $A_j = (\frac{\partial F}{\partial u_{j-s}}, \dots, \frac{\partial F}{\partial u_{j+s}})_{u_k=u_k^*}$. If the operator $\mathcal{L} - \lambda I$ has a bounded inverse one then $\lambda \notin \text{spec}\mathcal{L}$. Let us express this fact in a finite dimensional fashion. Because of Assumption 3.1, the equation

$$A_j\left(\{\xi_j\}^s\right) - \lambda\xi_j = 0 \tag{18}$$

can be rewritten as follows:

$$\xi_{j+s} = C_j^\lambda\left(\xi_{j-s}, \dots, \xi_{j+s-1}\right) \tag{19}$$

where C_j^λ is a linear form.

Let $\xi_{j-s} = \eta_1^j, \dots, \xi_{j+s-1} = \eta_{2s}^j$ then the latter equation is equivalent to the system $\eta^{j+1} = B_0^\lambda(x_j)\eta^j$, $\eta^j = (\eta_1^j, \dots, \eta_{2s}^j)$, or

$$\eta_1^{j+1} = \eta_2^j, \quad \dots, \quad \eta_{2s-1}^{j+1} = \eta_{2s}^j, \quad \eta_{2s}^{j+1} = C_j^\lambda(\eta_1^j, \dots, \eta_{2s}^j)\,. \tag{20}$$

In other words, for the orbit $\{x^j\} = \chi^{-1}(u^*)$ the cocycle $B_m^\lambda(x^j)$ is well-defined. Notice that $B_0^1(x^j) = D\,G(x^j)$, the differential of the map G.

Let us remind (see for instance [40] and references therein) that the cocycle is the sequence of linear maps $B_m^\lambda(x^j) : \mathbb{R}^{2s} \to \mathbb{R}^{2s}$ such that, for any $m > 0$,

$$B_m^\lambda(x^j) = B_0^\lambda(x^{j+m-1}) \circ \dots \circ B_0^\lambda(x^{j+1}) \circ B_0^\lambda(x^j)$$

and

$$B_{-m}^\lambda(x^j) = \left(B_0^\lambda(x^{j-m}) \circ \dots \circ B_0^\lambda(x^{j-2}) \circ B_0^\lambda(x^{j-1})\right)^{-1}\,.$$

The cocycle $B_m^\lambda(x^j)$ is said to be hyperbolic if

- There is a decomposition $\mathbb{R}^{2s} = E^s(x^j) \oplus E^u(x^j)$ such that $B_0^\lambda(x^j)\,E^s(x^j) = E^s(x^{j+1})$ and $B_0^\lambda(x^j)\,E^u(x^j) = E^u(x^{j+1})$.
- There are constants $c > 0$, $0 < \rho < 1$, $\gamma > 0$ such that if $v \in E^s(x^j)$ then for any $j \in \mathbb{Z}, m \in \mathbb{Z}_+$,

$$|B_0^\lambda(x^{j+m})| \le c\rho^m|v|$$

and if $v \in E^u(x^j)$ then

$$|B_0^\lambda(x^{j-m})| \le c\rho^m|v|\,.$$

- The angle $\measuredangle\left(E^s(x^j), E^u(x^j)\right) \ge \gamma$.

The following theorem explains why we remind these definitions.

Theorem 3.2. *If the operator $\mathcal{L}$ is self-adjoint and the cocycle $B_m^\lambda(x^j)$ is hyperbolic then $\lambda \notin$ spec $\mathcal{L}$.*

So, such values of λ do not belong to the spectrum of the operator and have nothing to do with stability of u^*. For example, if $\{x_j\}$ is hyperbolic

then, because of $B_1^0(x^j) = D\,G(x^j)$, $\lambda = 1$ does not belong to spec $\mathcal{L}$, and this orbit might correspond to a stable fixed point.

The theorem can be proven by straightforward construction of the Green function $(\mathcal{L} - \lambda I)^{-1}$ and by showing that this operator is bounded in the l^∞-norm (L. Glebsky, unpublished). Let us remark that if $\mathcal{L}$ is not self-adjoint, one can consider $\mathcal{L}^*\mathcal{L}$ and apply the theorem.

Of course it is difficult to check conditions of hyperbolicity of specific cocycles, but sometimes it is possible to do so.

Example 3.2. Coming back to the (4), for all j let $u_j^* = u_0$, $f(u_0) = 0$, and let $f'(u_0) = a$. Then $D\,G = \left[\begin{smallmatrix} 0 & 1 \\ -1 & 2-\frac{\alpha a}{\epsilon} \end{smallmatrix}\right]$, and $B_0^\lambda(x_j) = \left[\begin{smallmatrix} 0 & 1 \\ -1 & 2+\frac{\lambda-\alpha a-1}{\epsilon} \end{smallmatrix}\right]$, the constant matrices. Therefore hyperbolicity means the existence of the eigenvalues $\mu_{1,2}$ such that one of them is greater than one in modulus (since $\mu_1\mu_2 = 1$ another one will be less than 1). The characteristic equation for $B_0^\lambda(x_j)$ is $\mu^2 - \mu b + 1 = 0$, where $b = 2 + \frac{\lambda-\alpha a-1}{\epsilon}$. Those λ for which $|\mu_{1,2}| \neq 1$, do not belong to spec $\mathcal{L}$, therefore we are only interested in $\mu_{1,2} = e^{\pm i\phi}$.

It is simple to see that for every b, $|b| \leq 2$, the characteristic equation has a solution $\mu = e^{i\phi}$. The inequality $|b| \leq 2$ can be rewritten as $-4 \leq \frac{\lambda-\alpha a-1}{\epsilon} \leq 0$ or

$$1 + \alpha a - 4\epsilon \leq \lambda \leq 1\alpha\ . \tag{21}$$

Thus, if

$$4\epsilon - 2 < \alpha\, a < 0 \tag{22}$$

then the absolute value of every point in the spectrum of $\mathcal{L}$ is less than 1 and u^* is stable in the l^∞-norm.

In this example the configuration u^* is spatially homogeneous and the spectrum of $\mathcal{L}$ can be computed directely. Indeed, in this case this operator becomes

$$(\mathcal{L}\xi)_j = (1 + \alpha a - 2\epsilon)\xi_j + \epsilon(\xi_{j-1} + \xi_{j+1})\ .$$

A direct calculation shows that for every $\omega \in [0, 1)$, the configurations $\xi_j = \cos(2\pi\omega j)$ and $\xi_j = \sin(2\pi\omega j)$ (Fourier modes) are eigenvectors in l^∞ with eigenvalue

$$\lambda(\omega) = 1 + \alpha a - 2\epsilon + 2\epsilon\cos(2\pi\omega)\ .$$

The relation (21) immediately follows. A standard functional analysis argument shows that the spectrum is the same interval $[1 + \alpha a - 2\epsilon, 1 + \alpha a]$ in l^2.

While the argument applied here use translation invariance of the configuration u^*, the method of cocycles does not. So it could extend in principle to spatially periodic (and non-periodic) fixed points.

3.3 Approximation by Finite Dimensional Operators

The idea is to approximate the spectrum of the linearization operator $\mathcal{L}$ by spectra of finite dimensional matrices. We have a linear map $\xi' = \mathcal{L}\xi$. Given $N \in \mathbb{N}$, $N > s$, let the operator $\mathcal{L}_N : \mathcal{B} \to \mathbb{R}^{(2N+1)^d p}$ be defined as follows:

$$\begin{gathered}(\mathcal{L}_N\xi)_j = 0, \quad |j| > N,\\ (\mathcal{L}_N\xi)_j = (\mathcal{L}\xi)_j \quad |j| < N - s\\ (\mathcal{L}_N\xi)_j \text{ is an arbitrary bounded linear form if } \quad N - s \le |j| \le N\,.\end{gathered}$$

Such an operator can be treated as the map of $\mathcal{B}$ into itself.

Theorem 3.3. *[9, 15] Assume that $\mathcal{B}$ is endowed with the l_q^2-norm, $q \ge 1$. If $\mathcal{L}$ is normal and all $\mathcal{L}_N$ are normal for $N \ge N_0$ then spec $\mathcal{L} \cup \{0\} \subset$* clos $\left(\bigcup_{N=n_0}^{\infty} \textit{spec}\,\mathcal{L}_N\right)$.

By using this theorem one can find conditions of linear stability of specific fixed points and other orbits of CML (see for example [15, 16, 17]). In order to use this theorem one has to estimate the spectrum of the matrices $\mathcal{L}_N$. One can do it by using Gershgorin disks [25], since it is known that for any spectral value of λ of the matrix (a_{ij}) the inequalities

$$|\lambda - a_{ii}| \le \sum_{k \ne i} |a_{ij}|$$

hold. For the example (4) this inequality becomes the conditions (21) and we get (22).

Unfortunately, for l_q^2-norm, $q > 1$, linear stability does not imply stability with respect to the original non-linear system. Some discussions about the l_q^2-norm can be found in [18].

3.4 Spatial Chaos

It is well known from experimental works that disordered spatial field distributions (temperature, velocity, etc) occur in unbounded media with dissipation and energy pumping (see for instance [19]). If such a medium is described in the form of a LDS then the picture of spatial disorder can be expressed as follows.

Definition 3.1. *[20] A LDS is said to have spatial chaos if there exists a TDS-invariant set Λ of stable steady-state configurations on which TDS manifests a chaotic behavior (for example, the topological entropy of TDS is positive).*

Consider again the system (4) with $f(z) = z(z-c)(1-z)$, $0 < c < 1$, and $\alpha > 0$. In the uncoupled situation ($\epsilon = 0$) the system (4) has infinitely many steady-states solutions $u^* = (\ldots, u^*_{-1}, u^*_0, \ldots)$, $u^*_j \in \{0, c, 1\}$, $j \in \mathbb{Z}^d$. They form a TDS-invariant set Λ_3, and $\mathrm{TDS}|_{\Lambda_3}$ is topologically conjugated to the full (Bernoulli) shift with three symbols in the l^2_q-norm for an arbitrary $q > 1$. Furthermore, $D\mathcal{F}|_{u^*} = \mathrm{diag}\ \left(1 + f'(u^*_j)\right)$, where

$$f'(u^*_j) = \begin{cases} -\alpha c, & u^*_j = 1 \\ -\alpha(1-c), & u^*_j = 0 \\ \alpha(1-c)c, & u^*_j = c\,. \end{cases}$$

It follows that every solution u^*, where $u^*_j \in \{0; 1\}$, is stable in the l^2_q-norm, $q \geq 1$, provided that $1 - \alpha\, c > -1$, $1 - \alpha(1-c) > -1$. Thus, the uncoupled system has spatial chaos since $\mathrm{TDS}|_{\Lambda_2}$ is conjugated (for $q > 1$) to the full shift with two symbols (so, the topological entropy is $\log 2$), where $\Lambda_2 \subset \Lambda_3$ is the set of all configurations $\{u^*_j\}$, $u^*_j \in \{0; 1\}$. Notice that the solution $\{u^*_j\} \in \Lambda_3$ for which at least one $u^*_{j_0} = c$ is unstable.

If $0 < \epsilon \ll 1$ then, instead of Λ_2, we will have a TDS-invariant set $\Lambda_2(\epsilon)$ consisting of stable steady-state solutions close to the corresponding solutions in Λ_2, so the system (4) has spatial chaos either.

In the same way, one can show [15] that the system (5) with the same nonlinearity $f(z)$ possesses spatial chaos. Under the same conditions it has a TDS-invariant set $\Lambda^{(2)}_2(\epsilon)$ labeled by infinite matrices (a_{ij}), $(i, j) \in \mathbb{Z}^2$, $a_j \in \{0; 1\}$ (i.e., the elements of the full $\mathbb{Z}^2$-shift symbols), and the action of TDS on $\Lambda^{(2)}_2(\epsilon)$ is topologically conjugated to the full $\mathbb{Z}^2$-shift with two symbols in the l^2_q-norm for any fixed $q > 1$. Each of these configurations is asymptotically stable in the l^2_q-norm, $q > 1$.

Let us remark that for systems with spatial chaos it is natural to expect that the maximal attractors are infinite dimensional; in fact, it is generally true for a LDS (see for instance [21] and references therein). Therefore people use the Kolmogorov ϵ-entropy to characterize their complexity.

3.5 Homoclinic Points of TDS

It is possible to work on the problem of spatial chaos when a LDS is described by simple formulas as above. But in more complex situations one needs to have some verifiable criteria. For example, for a smooth dynamical system, the existence of transversal homoclinic trajectory implies the existence of Smale horseshoe. Thus, transversal homoclinic orbits serve as indicators of some chaotic behavior. As we shall see now homoclinic orbits of TDS may also serve as indicators of spatial chaos provided that some additional conditions hold.

Definition 3.2. *[20] A steady state solution u is said to be a homoclinic point of TDS if there exists $u_0 \in \mathbb{R}^p$ such that $\lim_{|j|\to\infty} u_j = u_0$.*

It follows from the continuity of $\mathcal{F}$ that the constant configuration $u^0 = \{u_j^0 = u_0,\ \forall j\}$ is also a solution of (8). We will deal here with the l^∞-norm (in [20] we considered the l^2-norm). Let $\sigma\left(D\mathcal{F}(u)\right)$ be the spectrum of the linearization operator $D\mathcal{F}$ at a fixed point u.

Definition 3.3. *1. The fixed point u is asymptotically stable if $\sigma\left(D\mathcal{F}(u)\right) \subset \{\lambda \,|\, |\lambda| < 1\}$ and hyperbolic if $\sigma\left(D\mathcal{F}(u)\right) \cap \{\lambda \,|\, |\lambda| = 1\} = \emptyset$.*
2. If u is hyperbolic then the dimension of $p\mathcal{B}$ is denoted by index $\left(D\mathcal{F}(u)\right)$ where $p\mathcal{B}$ is the projection of $\mathcal{B}$ associated to the spectrum subset $\sigma\left(D\mathcal{F}(u)\right) \cap \{\lambda \,|\, |\lambda| > 1\}$.

Let us remark that:

- If u is the steady state solution then $S_j u$, $j \in \mathbb{Z}^d$ is also a steady state solution,
- the configuration u is stable in the l^∞-norm iff it is stable in the l^2-norm [22].
- If u is a homoclinic point for the constant fixed point u^0 and u^0 is stable then $\sigma\left(D\mathcal{F}(u)\right) \cap \{\lambda \,|\, |\lambda| > 1\}$ consists of finitely many isolated eigenvalues with finite multiplicity.

By using these facts it was shown in [22] that if there exists a homoclinic solution u then there are infinitely many "multi-hump" homoclinic solutions. More precisely, we proved

Theorem 3.4. *[20, 22] Suppose that u is a homoclinic point for the constant fixed point u^0, u^0 stable and u is hyperbolic in the l^∞-norm. Then there exists $r > 0$ such that for every finite or infinite collection of indices $\{i_k\} \in \mathbb{Z}^d$ satisfying $|i_m - i_n| \geq r$, $m \neq n$,*

1. there exists a steady state ("multi-hump") solution $u(\{i_k\})$ that is close to the union $\bigcup_{j\in\{i_k\}} S_j u$ (a more detailed formulation is found in [20, 22]);
2. if u is stable then $u(\{i_k\})$ is stable either.
3. if u is unstable then index $\left(D\mathcal{F}(u(\{i_k\})\right) = card\ (\{i_k\}) \cdot index\ \left(D\mathcal{F}(u)\right)$ where card means the cardinality of a set.

Let us remark first that for the existence of multi-hump solutions the assumption of hyperbolicity is not necessary, it is sufficient that $\sigma\left(D\mathcal{F}(u)\right) \not\ni \{1\}$ (see [12]). Second, homoclinic orbits of TDS can be treated as defects (localized patterns) of the field, and the number r becomes an important parameter of the field that could be estimated by using quantities related to the LDS and to the solution u [23].

3.6 Heteroclinic Points of TDS

Let $u^1 = \{u_j^1\} = u_0^1$, $u^2 = \{u_j^2\} = u_0^2$, $j \in \mathbb{Z}^d$, be two spatially-homogeneous fixed points of a LDS.

Definition 3.4. *[22] A steady-state solution $u^{12} = \{u_j^{12}\}$ is said to be a heteroclinic point for the pair u^1, u^2 in the j_0–direction, $j_0 \in \mathbb{Z}^d$, if for any $i \in \mathbb{Z}^d$,*

$$|u_{tj_0+i} - u_0^1| \to 0,\ t \to -\infty,\ \text{and}\ |u_{tj_0+i} - u_0^2| \to 0,\ t \to \infty\ .$$

If $u^1 = u^2$ then u^{12} is called a homoclinic point in the j_0–direction.

Trivially, for $d = 1$ one has only one direction.

There is an interesting analogy between heteroclinic points of smooth dynamical systems and the ones of TDS. We describe two results.

Closed Heteroclinic Contours

Assume that for the pair of spatially–homogeneous points u^1 and u^2 there are heteroclinic points u^{12} and u^{21}.

Theorem 3.5. *[22]*

1. *If u^1, u^2, u^{12} and u^{21} are all linearly stable in the l^∞-norm then there exist stable homoclinic points in the j_0–direction of TDS for u^1 as well as for u^2.*
2. *If u^1, u^2 are stable and u^{12}, u^{21} are hyperbolic then there are hyperbolic homoclinic points in the j_0–direction for u^1 as well as for u^2.*

Let us remark that this theorem is similar in a sense to the following well-known result of the smooth dynamical systems theory.

Let u^1 and u^2 hyperbolic periodic points and their stable and unstable manifolds intersect transversally, i.e., $W^u_{u^1} \cap W^s_{u^2} \neq \emptyset$, $W^s_{u^1} \cap W^u_{u^2} \neq \emptyset$, then there are "transversal" homoclinic orbits for u^1 and u^2. In the theorem conditions of the transversality of intersections are replaced by conditions of stability (or hyperbolicity).

One can ask, say for $d = 1$, if is it true that the hyperbolicity of $u^1,\ u^2,\ u^{12}$ and u^{21} with respect to the evolution operator implies the "transversality" of intersections of stable and unstable sets (which are defined if one considers the TDS) of u^1, u^2 along u^{21}, u^{12}, or vice versa. We do not know the answer, and the main obstacle is the difficulty of right definitions of smoothness and transversality in the l^∞–norm of the action of the TDS on the set of steady-state solutions.

Classification of Heteroclinic Orbits in Morse–Smale Systems and Heteroclinic Points of TDS

A Morse–Smale system has only finitely many periodic orbits but may have infinitely many heteroclinic ones. To describe them one can use the partial order introduced in [30] and the notion of "beh" introduced in [31]. Let us remind them. We will speak about periodic orbits of the smooth dynamical system but keep in mind that they will be replaced below by spatially–homogeneous solutions of TDS.

One writes $u^1 \le u^2$ if $W^s_{u^2} \cap W^u_{u^1} \neq \emptyset$, i.e. there is a heteroclinic orbit from u^1 to u^2. Let us remind that for Morse–Smale systems all intersections are transversal. If $u^1 \le u^2$ then there is an ordered collection $\{v^1, \dots, v^m\}$ of periodic orbits such that $u^1 \le v^1 \le v^2 \le \dots \le v^m \le u^2$ and every intersection $W^u_{u^1} \cap W^s_{v^1}, W^u_{v^1} \cap W^s_{v^2}, \dots, W^u_{v^m} \cap W^s_{u^2}$ contains the only finitely many heteroclinic orbits. In this case the orbits belonging to $W^u_{u^1} \cap W^s_{u^2}$ are said to have beh equals m. If $m > 0$ then this intersection contains infinitely many heteroclinic orbits and they can be described in terms of a topological Markov chain [32].

For TDS, one again may say that $u^1 \le u^2$ if there is a heteroclinic point u^{12} of TDS. We proved in [22] that if $u^1 \le u^2$ and $u^2 \le u^3$ then $u^1 \le u^3$ and there exist infinitely many heteroclinic orbits joining u^1 and u^3. But we did not make a classification of heteroclinic orbits similar to the one described above. It is a nice open mathematical problem even for $d = 1$.

4 Spatially-homogeneous Solutions

Spatially-homogeneous solutions $v(n)$ satisfy the (9) and, by definition, belong to $\mathcal{B}$ provided that $|v(n)| \le \text{const} < \infty$.

4.1 Stability in the "Nearest Neighbors" Case

Let us consider $d = 1$ and the LDS

$$u_j(n+1) = F(u_{j-1}(n), u_j(n), u_{j+1}(n)) \,. \tag{23}$$

Any spatially-homogeneous solution $u_j(n) = v(n)$ satisfies the equation

$$v(n+1) = F(v(n), v(n), v(n)) \,.$$

Assume that $|v(n)| \le \text{const} < \infty$. The linearized equation for perturbations is $\xi(n+1) = \mathcal{L}(n)\xi(n)$ or

$$(\mathcal{L}(n)\xi(n))_j = A_1(n)\xi_{j-1}(n) + A_2(n)\xi_j(n) + A_3(n)\xi_{j+1}(n) \tag{24}$$

where $A_1(n) = F'_{u_{j-1}}(v(n), v(n), v(n))$, $A_2(n) = F'_{u_j}(v(n), v(n), v(n))$, $A_3(n) = F'_{u_{j+1}}(v(n), v(n), v(n))$. By using Gershgorin disks, one can prove

that if $A_1(n) = A_3(n)$ and $A_2(n) - 2|A_1(n)| > -1+\delta$, $A_2(n)+2|A_1(n)| < 1-\delta$, $0 < \delta < 1$, then $\{v(n)\}$ is stable in the l^2-norm.

One can use this approach to find conditions of stability of spatially-homogeneous solutions in LDS with $d > 1$ and more general couplings. A more sophisticated approach to study stability of spatially-homogeneous solutions was proposed in [14].

4.2 Unidirectional Coupling

For $d = 1$, there is a class of models reflecting the development of perturbations along physical coordinate [26]. Some of them may be described as semi-infinite LDS of the form [27]:

$$u_j = f(u_j(n)),\ 0 \le j \le k \tag{25}$$

$$u_j = f(u_j(n)) - \gamma \left[\phi(u_j(n)) - \sum_{s=1}^{r} \alpha_s [\phi(u_{j-s}(n))] \right] + \epsilon \sum_{t=-l}^{l} \beta_t u_{j+t}(n),\ r, l \le k\ . \tag{26}$$

We proved in [27] that if the system (25) has a hyperbolic attractor, then there exists $q > 1$ such that the system (25, 26) has a finite-dimensional hyperbolic (in the l^2_q-norm) attractor if ϵ, γ are small enough and some other conditions are satisfied. It does not imply, however, that for the l^∞-norm the dimension of the attractor could not grow along spatial coordinate and could not be infinite for the system (25, 26).

4.3 A Particular Case

We consider the system

$$x_j(n+1) = f(x_j(n)) - \gamma\psi(x_{j-1}(n), x_j(n)), \quad x_j \in \mathbb{R},\ j > 0 \tag{27}$$

with the boundary condition $x_0(n+1) = x_0(n)$ where f, ψ are smooth functions and $\psi(x, x) = 0$. If we assume that $x_j(0) = x_0(0)$, then $x_j(n) = x_0(n)$, $j \in \mathbb{Z}_+$, will be a spatially-homogeneous solution of (27). We assume that $|x_0(n)| \le \text{const} < \infty$. According to the linearized system one can single out two types of coupling. The linearized equation has the form

$$\xi_j(n+1) = \left(f'(x_0(n)) - \gamma \frac{\partial \psi}{\partial x_j}(x_0(n), x_0(n)) \right) \xi_j(n) - \gamma \frac{\partial \psi}{\partial x_{j-1}}(x_0(n), x_0(n))\ \xi_{j-1}(n)\ . \tag{28}$$

We introduce the Lyapunov exponent for a function $w(n)$ as follows

$$\lambda(w(n)) = \overline{\lim}_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} w(n).$$

Let $\lambda_\gamma = \lambda(f'(x_0(n)) - \gamma \frac{\partial \psi}{\partial x_j}(x_0(n), x_0(n)))$, so that $\lambda_0 = \lambda\left(f'(x_0(n))\right)$. Thus, one can treat λ_γ as the Lyapunov exponent of the cocycle on the orbit $\{x_0(n)\}$.

Definition 4.1. *The coupling is said to be* dissipative *if* $\lambda_\gamma < \lambda_0$ *and* inertial *if* $\lambda_\gamma = \lambda_0$.

The simplest example of dissipatively coupled maps is the CML

$$x_j(n+1) = f(x_j(n)) - \gamma\left(f(x_j(n)) - f(x_{j-1}(n))\right) . \tag{29}$$

An interesting problem is to find an example of inertially coupled LDS. One may assume that in the stable situation, generally, any coupling is a combination of dissipative and inertial couplings and one can believe that it is so for the system

$$x_j(n+1) = f(x_j(n)) - \gamma\left(x_j(n) - x_{j-1}(n)\right) . \tag{30}$$

For the system (29) it is possible to derive conditions of stability in terms of the Lyapunov exponent λ_0 of the individual map. Let us remark that, for any $\epsilon > 0$ there exists $N(\epsilon)$ such that for $N > N(\epsilon)$ the following estimate holds

$$\prod_{k=m}^{m+N-1} |f'(x_0(k))| \le C_m e^{(\lambda_0+\epsilon)N}.$$

If we assume that the orbit $\{x_0(n)\}$ is the uniformly hyperbolic one for the individual map so that $C_m \le c < \infty$ then we arrive to the following theorem [28]

Theorem 4.1. *If*

$$\lambda_0 < -\ln|1-\gamma| - \frac{\gamma}{1-\gamma} , \quad 0 < \gamma < 1 \tag{31}$$

then $\xi_j(n) \to 0$ *uniformly in* j*, i.e.,* $\{x_0(n)\}$ *is linearly asymptotically stable in the* l^∞*-norm.*

The inequality (31) implies that $\lambda_0 < 0$, i.e., generally, $x_0(n)$ is a stable, eventually periodic, orbit.

4.4 Synchronization of Chaotic Motions in Infinite Lattices

One can say not too much about this problem. First of all, there is no rigorous results confirming the possibility to synchronize chaotic "oscillators" for infinite lattices. Numerical simulations by definition deal with finite lattices. Nevertheless, people who perform such simulations believe that for identical individual subsystems in chaotic regimes, subjected to some dissipative coupling, synchronized motions can occur. For example, the authors of [33] studied the system

$$x_j(n+1) = (1-\gamma_1-\gamma_2)f(x_j(n)) + \gamma_1 f(x_{j-1}(n)) + \gamma_2 f(x_{j+1}(n)) \qquad (32)$$

$j = 1, 2, \ldots, N$, with boundary conditions

$$x_1(n+1) = (1-\gamma_2)f(x_1(n)) + \gamma_2 f(x_2(n))$$

$$x_N(n+1) = (1-\gamma_1)f(x_N(n)) + \gamma_1 f(x_{N-1}(n))$$

where $f(x) = \epsilon - x^2$. It has been shown that for $\epsilon = 1.67$ the individual system $x_j(n+1) = f(x_j(n))$ manifests a chaotic behavior and for $\gamma_1 = 0.7, \gamma_2 = 0.1$ the diagonal $x_1 = x_2 = \ldots = x_N$ is stable in transversal directions even for $N \gg 1$.

So, an interesting open problem now is to prove that chaotic synchronization can (or can not) occur in infinite lattices of identical elements, to describe conditions on individual systems, couplings, etc. For specific cases it was shown in [29] and (Yu. Maistrenko, unpublished) that synchronization regimes do not exist in infinite LDS. However a question arises: even if such regimes exist how robust can they be with respect to small perturbations of individual subsystems.

Let us try to formulate an intuitive scenario. Consider the system

$$x_j(n+1) = f(x_j(n), \mu_j) + \gamma F(\{x_j(n)\}^s) \qquad (33)$$

$0 \le j \le N$, where the individual "rule" f depends on a parameter μ, and the displacement of parameters $\delta = \sup_{j,j'} |\mu_j - \mu_{j'}|$ is very small, i.e. $\delta \ll 1$. Assume also that for equal values of parameters $\mu_j = \mu_0$ the system (33) possesses a synchronized chaotic regime for $\gamma = \gamma_0$. There is a logical possibility that the size of the region of values of γ for which a chaotic synchronized regime occurs behaves asymptotically as $N^{-\beta}$ where $\beta > 0$ depends on d, on boundary conditions and on the function f. Thus, the problem is to find conditions of chaotic synchronization and to show that such regimes can (or can not) be robust.

5 Traveling Waves

As far as in the PDE-theory a stationary traveling wave solution depends on a linear combination of spatial and temporal variables, $u(x,t) = \psi(lx + vt)$,

where l is the wave vector, lx is the scalar product and v is the velocity. For LDS this definition works as well and one can define

$$u_j(n) = \psi(lj + mn) \tag{34}$$

as a traveling wave solution. One can immediately see from this definition that fixed points of operator $S_{-j_0} \circ \mathcal{F}^k$ are traveling wave solutions of the type (34) if $lj_0 = mk$.

Let us remark that some methods and ideas presented here are similar to the ones from the Sect. 3. Nevertheless, there are differences because both the evolution and translation operators act on the set of traveling waves non-trivially. This fact implies different dimensions and norms for different embeddings, and therefore a repetition will be not completely unjustified.

5.1 Finite-Dimensional Description

Substituting (34) in (3) we obtain the following expression

$$\psi(lj + mn + m) = F(\{\psi(lj + li + mn)\}_{|i|\le s}) . \tag{35}$$

We assume that

$$\max_{|i|\le s} |li| < m \tag{36}$$

then (35) can be rewritten in a finite dimensional form. Indeed, let us order the collection of numbers $\{li\}_{|i|\le s}$ in the natural way $\{li\}_{|i|\le s} = \{p_1, p_2, \dots, p_{N(s)}\}$, $p_k < p_{k+1}$; some of these values can have multiplicity greater than 1. Then (35) can be presented as follows:

$$\psi(lj + mn + m) = G(\psi(lj + p_1 + mn), \dots, \psi(lj + p_{N(s)} + mn)) . \tag{37}$$

Sometimes (37) is called the traveling wave equation. Because of the assumption (36), $p_1 > -m, p_{N(s)} < m$. We introduce the "traveling coordinate" $t = lj + mn + m$, then (37) becomes

$$\psi(t) = G(\psi(t - m + p_1), \dots, \psi(t - m + p_{N(s)})) . \tag{38}$$

This equation can be rewritten in the form of a dynamical system generated by a map $T : x \to \overline{x}$ where $x = (x_1, \dots, x_M)$, $x_k = \psi(lj + mn + p_1 - 1 + k)$, $k = 1, \dots, M$, $M = m - p_1$ and T is expressed as follows

$$\overline{x}_k = x_{k+1}, k = 1, \dots, M - 1, \quad \overline{x}_M = G(x_1, x_{p_2 - p_1}, \dots, x_{p_{N(s)} - p_1}) . \tag{39}$$

Example 5.1. We consider again the system (4) (so $d = 1$) and choose $m = 3, l = 2$, so that $M = 3 - (-2) = 5$, $u_j(n) = \psi(2j + 3n)$. The (38) becomes

$$\psi(t) = \psi(t - 3) + \alpha f(\psi(t - 3)) + \epsilon(\psi(t - 5) - 2\psi(t - 3) + \psi(t - 1)) \tag{40}$$

and if we introduce the notations: $x_1 = \psi(t-5), x_2 = \psi(t-4), x_3 = \psi(t-3), x_4 = \psi(t-2), x_5 = \psi(t-1)$, then the map $T : \mathbb{R}^5 \to \mathbb{R}^5$ looks as follows

$$\overline{x}_1 = x_2, \ldots, \overline{x}_4 = x_5, \quad \overline{x}_5 = x_3 + \alpha f(x_3) + \epsilon(x_1 - 2x_3 + x_5) \, . \tag{41}$$

Every bounded orbit of the dynamical system (41) corresponds to a solution of the traveling (40) and therefore, to a traveling wave solution of the original system.

Following the work [34], we describe a construction of this correspondence for the case $d = 1$.

5.2 Two Embeddings

For $d = 1$ the traveling wave (38) becomes

$$\psi(t) = F(\psi(t - m - ls), \psi(t - m - l(s-1)), \ldots, \psi(t - m + ls)) \tag{42}$$

where t is the traveling coordinate, and the assumption (36) becomes

$$m \geq ls + 1 \, . \tag{43}$$

Let $x_1(t) = \psi(t - m - ls), x_2(t) = \psi(t - m - ls + 1), \ldots, x_{ls+1}(t) = \psi(t - m), x_{ls+m}(t) = \psi(t - 1)$, then the dynamical system generated by the map G can be written as

$$\begin{aligned} x_1(t+1) = x_2(t), \ldots, x_{ls+m-1}(t+1) = x_{ls+m}(t), \\ x_{ls+m}(t+1) = F(x_1(t), x_{1+l}(t), \ldots, x_{2ls+1}(t)) \, . \end{aligned} \tag{44}$$

It simple to see that if $\det \frac{\partial F}{\partial x_1} \neq 0$ then G is a diffeomorphism. We assume that this condition holds.

The orbits of the map $G : \mathbb{R}^{p(ls+m)} \to \mathbb{R}^{p(ls+m)}$ can be embedded into the phase space $\mathcal{B}$ of LDS. One of these embeddings can be defined as follows. Let the map $\chi : \mathbb{R}^{p(ls+m)} \to (\mathbb{R}^p)^{\mathbb{Z}}$ be defined by formulas

$$\chi(x)_j = \begin{cases} x_j & \text{if} \quad -ls - m \leq j \leq -1 \\ (G^{i+1}x)_{ls+m} & \text{if} \quad j > -1 \\ (G^{j+ls+m}x)_1 & \text{if} \quad j \leq -ls - m - 1 \, . \end{cases}$$

The following theorem holds:

Theorem 5.1. *[34]. There exists $q_0 \geq 1$ such that for any $q > q_0$ we have*

1. $\chi(\mathbb{R}^{p(ls+m)}) \subset \mathcal{B}$, where $\mathcal{B}$ is endowed with the l^2_q-norm,
2. χ is a smooth injective map, and
3. $S_1 \circ \chi = \chi \circ G$, the map G commutes with the shift S_1 generating the TDS.

For the l^2_q–norm, $q < q_0$, or for the l^∞–norm not all orbits of G are bounded and the map χ is not defined for all points in $\mathbb{R}^{p(ls+m)}$. Nevertheless the restriction of χ of the set Λ of all points of all bounded orbits of G is well-defined and still satisfies the last statement of the theorem.

We describe now how the evolution operator acts on the set of traveling wave solutions. Let $\psi \in \chi(\mathbb{R}^{p(ls+m)})$ and define the map α by $(\alpha\psi)_j = \psi(lj)$.

Theorem 5.2. *Under the conditions of Theorem 5.1 and for $q > q_0$ we have*

1. $\alpha(\mathbb{R}^{p(ls+m)}) \subset \mathcal{B}_{q^l}$
2. α is smooth map and
3. $\mathcal{F}^l \circ \alpha = \alpha \circ S^m$.

Corollary 5.1. *Let A be a hyperbolic locally maximal set of the map G then $\mathcal{A} = \alpha(\chi(A))$ is a hyperbolic locally maximal set of $\mathcal{F}^l$ in $\mathcal{B}_{q^l}$.*

In the l^2_q-norm, $q < q_0$ or in the l^∞-norm the map α is still well-defined on the set $\chi(\Lambda)$, and the last statement of Theorem 5.2 holds.

Thus the dynamics of the LDS on the set of traveling waves is defined, in fact, by the map G that has a special form. We will call G *the generalized Hénon-type map.*

5.3 Generalized Hénon-type Maps

Something can be said about dynamics of the map G in the case of weak couplings. We consider the system

$$u_j(n+1) = f(u_j(n)) + \epsilon T(u_{j-s}(n), \dots, u_{j+s}(n)) \; . \tag{45}$$

The map $G : X \to \overline{X}$, $x = (x_1, \dots, x_{ls+m})$, becomes

$$\begin{aligned} \overline{x}_1 = x_2, \dots, \overline{x}_{ls+m-1} = x_{ls+m} \\ \overline{x}_{ls+m} = f(x_{ls+1}) + \epsilon T(\{x_{p(i)} u_{j-s}\}_{i \le s}) \end{aligned} \tag{46}$$

where $x_k \in \mathbb{R}^p$, $p(i) = li + 1$, $i = 0, 1, \dots, s$. The following statement was proved in [34].

Lemma 5.1. *It $f : \mathbb{R}^p \to \mathbb{R}^p$ is a diffeomorphism possessing a hyperbolic locally maximal set then there exists $\epsilon_0 > 0$ such that for every ϵ, with $|\epsilon| < \epsilon_0$, the map G possesses a hyperbolic locally maximal set provided that $det \frac{\partial T}{\partial x_1} \neq 0$.*

What one can say if f is not one-to-one? Some results related to this problem can be found in [35]. We consider now an example of such a situation in the system (4). We will seek traveling wave solutions in the form $u_j(n) = \psi(j + n)$ [36]. The traveling wave equation can be written as

$$\psi(j+n+1) = \frac{1}{1-\epsilon}[(1-2\epsilon)\psi(j+n) + \epsilon\psi(j+n-1) + \alpha f(\psi(j+n))]$$

and if we set $x_n = \psi(j+n-1)$, $y_n = \psi(j+n)$ then we arrive to the system

$$x_{n+1} = y_n, \quad y_{n+1} = \frac{\epsilon}{1-\epsilon}x_n + \frac{1}{1-\epsilon}[(1-2\epsilon)y_n + \alpha f(y_n)]\,. \tag{47}$$

For example, if f is a cubic polynomial as in Subsect. 3.4 then one can find parameter values for which the system (47) has a Smale horseshoe [36], i.e., plenty of chaotic orbits exist. They correspond to traveling waves with "chaotic profiles" [36], see also [37]

Another related problem can be formulated as follows: Let f be a quadratic (or unimodal) map of the interval and $ls + m > 2$. Is it possible to generalize results of [38] to this case?

5.4 Abundance of Traveling Waves

We know that in a basic set of a smooth Axiom A system periodic points are dense, and this fundamental fact allow people to succeed in many difficult problems of hyperbolic dynamics (see for instance [40]). It would be not surprising if a similar statement would occur for LDS. But, as we study now traveling wave solutions, then the following conjecture seems to be interesting.

Conjecture 5.1. If Λ is closed locally-maximal transitive hyperbolic set of a LDS, then traveling waves solutions are dense in Λ.

Maybe one must assume also that Λ is compact. This conjecture is based on some results that we are going to describe now.

Coupled Extended Circle Maps [41]

Let S be unit circle, $S = \{x(\mathrm{mod}1)\}$, $S^{\mathbb{Z}} = \bigotimes_{i\in\mathbb{Z}} S_i$ where S_i are copies of S and $\mathbb{R}^{\mathbb{Z}} = \{\underline{x} = (\ldots, x_{-1}, x_0, x_1, \ldots)\}$. Let $\mathcal{B}_q = \{\underline{x} \in \mathbb{R}^{\mathbb{Z}}, \parallel \underline{x} \parallel < \infty\}$, $\parallel \cdot \parallel$ is the l_q^2-norm, $q > 1$. Denote by $\pi : \mathbb{R}^{\mathbb{Z}} \to S^{\mathbb{Z}}$ the projection: $\pi\,\underline{x} = (\ldots, x_{-1}(\mathrm{mod}1), x_0(\mathrm{mod}1), x_1(\mathrm{mod}1), \ldots)$, and endow $S^{\mathbb{Z}}$ with the metric

$$d(\pi\underline{x}, \pi\underline{y}) = \inf \parallel \underline{\tilde{x}} - \underline{\tilde{y}} \parallel$$

where the infimum is taken over all elements $\underline{\tilde{x}}, \underline{\tilde{y}} \in \mathcal{B}_q$ with $\pi\underline{\tilde{x}} = \pi\underline{x}$, $\pi\underline{\tilde{y}} = \pi\underline{y}$. We denote by $\mathbb{T}$ the set $S^{\mathbb{Z}}$ endowed with this metric. Let $f : S \to S$ be an expanding C^r-map, $r > 1$ and $F : \mathbb{R} \to \mathbb{R}$ its lift. We assume that $1 < F' < \infty$.

Now we introduce a LDS by

$$\left(\hat{\mathcal{F}}\underline{x}\right)_j = F(x_j) + \epsilon \sum_{i-j=-s}^{s} a_{i-j}(x_i)$$

where $a_k(x)$ are C^r-smooth 1-periodic functions. So, $\hat{\mathcal{F}}$ generates a LDS on $\mathcal{B}_q$ and $\mathcal{F}$ defined by

$$(\mathcal{F}\underline{x})_j = F(x_j) + \epsilon \sum_{i-j=-s}^{s} a_{i-j}(x_i), \bmod 1$$

generates a LDS on $\mathbb{T}$.

We call the solution of the form $x_j(n) = \psi(kj + mn)$ the traveling wave solution of the (k, m)-type.

Theorem 5.3. *There exists $\epsilon_0 > 0$ such that for each $\epsilon \in [0, \epsilon_0]$ and for every $k, m \in \mathbb{Z}$ the traveling wave solutions of the (k, m)-type are dense in $\mathbb{T}$.*

The theorem shows that one can approximate an arbitrary point on $\mathbb{T}$ by a point of a traveling wave solution of the (k, m)-type in such a way that the j-th coordinates of these configurations will be close to each other for $-N \le j \le N$, $N \gg 1$, but for other values of j the corresponding coordinates might be very different. In the following example it is not the case.

LDS with Hyperbolic Repellers in Individual Systems

We consider the system generated by the map (6) for which the individual map $f : I \to I$ of the interval is supposed to satisfy the following assumptions.

- There exists a finite collection of piecewise disjoint closed intervals $I_1, \ldots, I_N \subset I$,
- $|f'(x)| > 1$, $x \in I_i$, $i = 1, \ldots, N$,
- for any i there exists j such that $I_j \subset \mathrm{Int}\,(f(I_i))$.

These conditions imply that for the map f there exists a hyperbolic repeller $\Lambda_0 \subset \bigcup_{i=1}^N I_i$ consisting of all points of all orbits belonging to $\bigcup_{i=1}^N I_i$. The restriction $f|_{\Lambda_0}$ is topologically conjugated to the topological Markov chain (Ω_A, σ) where the transition matrix A is determined by the last assumption.

We proved in [43] that if the l^∞-norm $\| L - Id \| < \epsilon$, and ϵ is small enough, then the map $\mathcal{F}$ (generating the CML) has an invariant set $\tilde{\Lambda}$ such that $\mathcal{F}|_{\tilde{\Lambda}}$ is topologically conjugated to $\mathcal{F}_0|_{\tilde{\Lambda}_0}$ where $\mathcal{F}_0$ is the evolution operator for uncoupled system, $L = Id$, and $\tilde{\Lambda}_0 = \Lambda_0^{\mathbb{Z}}$.

It is useful now to rephrase the definition of traveling wave solutions. We say that u is a traveling wave configuration of the velocity $v \in \mathbb{Q}$ if $v = \frac{m}{l}$ and $\mathcal{F}^l u = S_1^m u$ where S_1 is the shift generating the TDS. It is simple to verify that a traveling wave configuration are points belonging to a traveling wave solutions of the form $\psi(lj + mn)$.

If $v \in \mathbb{R}\backslash\mathbb{Q}$, we say that u is a traveling wave configuration of the velocity v if for any sequence $\{(m_i, l_i)\}_{i\in\mathbb{N}}$, $m_i \in \mathbb{Z}$, $l_i \in \mathbb{N}$, such that $\lim_{i\to\infty} \frac{m_i}{l_i} = v$ one has

$$\lim_{i\to\infty} \| \mathcal{F}^{l_i} u - S_1^{m_i} u \| = 0.$$

Theorem 5.4. *If the topological Markov chain (Ω_A, σ) is mixing, $\| L - Id \| < \epsilon$, and ϵ is small enough then, for any $v \in \mathbb{R}$, the traveling wave configurations of the velocity v are dense in $\tilde{\Lambda}$ (in the l^∞-norm).*

Let us remark that in [44, 45, 46], the authors have proved the existence of traveling waves (the so-called fronts) with arbitrary real velocity and monotonic shape in bistable CML, see the chapter by R. Coutinho and B. Fernandez.

To finish this section let us mention that we omitted all results related to stability of traveling waves. Some of them can be found in [34, 37].

6 Weakly Coupled Systems

There are many results for weakly coupled LDS. In our framework those are systems of the form

$$u_j(n+1) = f(u_j(n), \mu_j) + \epsilon F\left(\{u_j(n)\}^s\right) \tag{48}$$

generated by the map that will be denoted by $\mathcal{F}_\epsilon$, where the coupling parameter ϵ is assumed to be small and μ_j is a parameter of the individual map. If $\mu_j = \mu_0$, $j \in \mathbb{Z}^d$ then the system (48) is of the type (3). A problem that naturally arises is to prove the topological similarity of coupled and uncoupled systems provided that the individual maps posses some hyperbolic properties. Such kind of structural stability, in fact, occurs, and it was proved by several authors in several situations. But in any of them it has been assumed that the individual map $f(x, \mu_j)$ has a hyperbolic locally-maximal set Λ_j with chaotic behavior of orbits. Then the product $\Lambda^0 = \bigotimes_{j\in\mathbb{Z}^d} \Lambda_j$ will be an invariant set for the uncoupled map $\mathcal{F}_0$ and will inherit some hyperbolic properties depending on the norm in $\mathcal{B}$. Furthermore, because of hyperbolicity, Λ_j can be described in terms of a topological Markov chain (Ω_{A_j}, σ_j) where A_j is a transition matrix. In other words, $f_j|_{\Lambda_j}$ is topologically semi-conjugated to the subshift (Ω_{A_j}, σ_j) (conjugated if Λ_j is a zero-dimensional repeller of f). Thus, the product (Σ, σ) where $\Sigma = \bigotimes_{j\in\mathbb{Z}} \Omega_{A_j}$, $\sigma = \bigotimes_{j\in\mathbb{Z}} \sigma_j$, gives a symbolic description of the uncoupled system. The hyperbolicity of $\mathcal{F}_0|_{\Lambda^0}$ allows us to prove the structural stability result by using sometimes the symbolic representation (Σ, σ). In particular, it was proved that for small values of ϵ there is a $\mathcal{F}_\epsilon$-invariant set Λ^ϵ. The following results are worth to be listed.

- The existence of the conjugacy between $\mathcal{F}_0|_{\Lambda^0}$ and $\mathcal{F}_\epsilon|_{\Lambda^\epsilon}$ which is a homeomorphism in the l^∞-norm for the spatially homogeneous situation ($\mu_j = \mu_0$) has been proved in [42] by means of Implicit Function Theorem (IFT). The authors considered short range interactions, i.e., F may depend on infinitely many variables but the dependence decays exponentially fast with the label of coordinates. Of course, the coupling in (48) automatically satisfies this assumption since F has only finitely many arguments.

- The existence of such a conjugacy again for the spatially homogeneous situation where the coupling of the convolution type was also proved in [43], as it was already mentioned, without using IFT, but only for the case where Λ_j is a zero-dimensional repeller on the interval. The proof is performed by using the symbolic system (Σ, σ), and it is more constructive in a sense that one can, in principle, control the magnitude of ϵ. Furthermore, the spatial dependence of interaction does not need to be of short range, just summability is sufficient.
- For homogeneous (Theorem 3.16 in [24]) and non-homogeneous (Theorem 7.1 in [24]) cases the existence of the conjugacy has been proved by using IFT again in some unusual norm (that is something in between the l^∞ – and the l^2_q-norm, $q > 1$).
- For the non-homogeneous case ($\mu_j \neq \mu_0$) the existence of conjugacy was proved in [48] for the l^2_q-norm, $q > 1$. Again the case of zero-dimensional repellers of the maps of the interval was considered and some additional natural conditions were assumed. The author used methods of [43, 49, 50] and did not use IFT.

  The conjugacy implies that all topological properties of a weakly coupled system are the same as for the uncoupled one. So, such important features of hyperbolic set as a proper symbolic description, the density of periodic points and traveling wave solutions in the transitive case, etc. are established without any difficulty. Furthermore, a calculation of some characteristics of chaos (for example the density of directional entropy) can be performed in a very simple way.

7 From Infinite to Finite Lattices. Concluding Remarks and Problems

Of course, there are many other interesting solutions in LDS which are worth to be studied and there are many other results about topological properties of some LDS. But, we believe that the ideas given in these lectures and their results will serve as a first step to be acquainted with this field. For results on ergodic theory of LDS see [47, 51] and references therein, and also some chapters in this volume.

We now describe some important problems when passing from infinite to finite lattices [10].

The system (4) is infinitely extended, i.e., $-\infty < j < \infty$. We now ask how the behavior of solutions of a large but finite lattice is related to the behavior of solutions of (4). We restrict ourself to the steady-state solutions and will impose different boundary conditions.

We consider solutions on a finite lattice, $-N \leq j \leq N$.

7.1 Periodic Boundary Conditions

Steady-state solutions of (4), i.e., solutions of (16) satisfying the identity $u_j \equiv u_{j+2N+1}$ can be treated as solutions on the finite lattice $-N \le j \le N$ with periodic boundary conditions.

We consider the simplest question: let P_N be the number of steady state solutions, how to describe the asymptotic behavior of P_N as $N \to \infty$? Of course, the question is valid only if $P_N \neq \infty$. If P_N is infinite for some N then the question could be reformulated: how many isolated steady-state solutions exist if $N \gg 1$?

Evidently, P_N equals the number of $(2N+1)$-periodic points of the system (16). We may specify the problem. Fix $R \gg 1$ and consider only periodic points belonging to the ball of radius R centered at the origin. Then we can expect that $P_N < \infty$ for any N. If so, then we can introduce the quantity

$$\lambda_P = \overline{\lim}_{N\to\infty} \frac{\ln P_N}{2N}, \quad \text{i.e.,} \quad P_N \sim e^{2\lambda_P N}.$$

So we have the expression of λ_P as the growth number of periodic points; very often such a growth number equals the topological entropy (see, for instance, [40]) and then

$$P_N \sim e^{2h_{top}N} . \tag{49}$$

7.2 Dirichlet Boundary Conditions

Assume that the desired solutions satisfy the relationship $u_{-N} = 0$, $u_N = 0$. This means that $x_{-N+1} = 0$, $y_N = 0$. In other words, any orbit segment (x_j, y_j) satisfying (16) for $-N+1 \le j \le N$ and the conditions

$$x_{-N+1} = 0, \qquad y_N = 0 \tag{50}$$

corresponds to a finite collection of points

$$u_{-N} = x_{-N+1} = 0, \quad u_{-N+1} = y_{-N+1}, \ldots, u_N = y_N = 0$$

which constitute a steady-state solution of (4) satisfying the Dirichlet boundary condition. Let us treat it in the following way. Consider the lines $L_1 = \{(x,y) : x = 0\}$, $L_2 = \{(x,y) : y = 0\}$ and let (x_N^*, y_N^*) be a point of intersection $G^{2N}L_1 \cap L_2$ where G is the map (16). Then the points $(x_{-N+1}^*, y_{-N+1}^*) = G^{-2N}(x_N^*, y_N^*)$, $(x_{-N+2}^*, y_{N+2}^*) = G^{-2N+1}(x_N^*, y_N^*) \cdots (x_{-N+j}^*, y_{-N+j}^*) = G^{-2N+j-1}(x_N^*, y_N^*), \ldots, (x_N^*, y_N^*)$ form the desired orbit segment.

In other words, the number of steady-state solutions satisfying the Dirichlet boundary conditions is equal to the number (say D_N) of points in the intersection $G^{2N}L_1 \cap L_2$.

If $\lambda_D = \overline{\lim}_{n\to\infty} \frac{\ln D_N}{2N}$, then we have an asymptotic relationship

$$D_N \sim e^{2\lambda_D \cdot N} . \tag{51}$$

7.3 Neumann Boundary Conditions

For a discrete spatial variable, the Neumann boundary condition can be expressed as follows: $u_{-N} = u_{-N+1}$, $u_{N-1} = u_N$, i.e.,

$$x_{-N+1} = y_{-N+1}, \quad x_N = y_N \ . \tag{52}$$

The number of steady-state solutions satisfying these conditions is equal to the number of the orbit segments (x_j, y_j), $-N+1 \le j \le N$, satisfying (52), and this number, say ξ_N, equals the number of points in the intersection $S^{2N}L \cap L$ where L is the diagonal $\{(x, y) : x = y\}$. Again we can introduce

$$\lambda_{Ne} = \overline{\lim}_{N\to\infty} \frac{\ln \xi_N}{2N}$$

so that

$$\xi_N \sim e^{2\lambda_{Ne} N} \ . \tag{53}$$

The problem of the asymptotic behavior of the number of points in the intersection $f^k L_1 \cap L_2$, where L_1, L_2 are submanifolds of a smooth manifold, and f is a smooth map (in particular a diffeomorphism) is said to be a problem of *dynamics of the intersection*. Such problems which have appeared recently in different branches of analysis are very interesting. There are some general results (see for instance [39] p. 261), but there are no approaches to solve specific problems. For example, the following open problems seem to be very attractive.

- Is it true that, in general

  $$\lambda_P = \lambda_D = \lambda_{Ne} \ ?$$

  If it is true for most nonlinearities $f(u)$ then we can introduce some kind of "structural stability" for infinitely extended systems. For such systems the number of steady-state solutions will be approximately the same for all boundary conditions if the size of the lattice $N \gg 1$. And, for example, in numerical simulations we may use arbitrary boundary conditions to study the solutions in an infinite lattice. If it is not true then the following problem appears.
- Which parameters are "responsible" for the values of λ_j, $j \in \{P, D, Ne\}$?
- What kind of bifurcations occurs if the line L_1 is moved so that its slope goes from $-\infty$ to ∞?

Acknowledgments

The author was partially supported by NSF-CONACyT grant E120933 and by CONACyT grant 485100-5-36445-E. He greatly appreciated J.-R. Chazottes and B. Fernandez for the opportunity to participate in CML2004 and for useful remarks about the text of this chapter.

References

1. H. Chaté, and M. Courbage, eds, *Lattice Systems*, Physisca D, **103** (1997). 153
2. I. Walter and R. Kapral, *Spatial and temporal structures in systems of coupled nonlinear oscillators*, Phys. Rev. A **30** (1994), pp 2047–2057. 153
3. K. Kaneko, *Period doubling and kink-antikink patterns, quasiperiodicity in antiferro-like structures and spatial intermittency in coupled logistic lattices*, Prog. Theor. Phys. **72** (1984), pp 480–497. 153
4. J.P. Crutchfield, *Space-time dynamics in video feedback*, Physica D **10** (1984), pp 229–233. 153
5. S.P. Kuznetzov and A.S. Pikovsky, *Universality of the period doubling bifurcation in one-dimensional dissipative media*, Izvestija VUZ, Radiofizika **28** (1985), pp 308–324. 153
6. K. Kaneko, ed., *Theory and application of Coupled Map Lattices* Wiley, Chichester, 1993. 153
7. L.A. Bunimovich and Ya.G. Sinai, *Space-time chaos in coupled map lattices* Nonlinearity **1** (1988), pp 491–516. 153
8. L.A. Bunimovich, *Lattice dynamical systems*, in: J.C. Robinson and P.A. Glendinning, eds., From Finite to Infinite-dimentional Dynamical Systems, Kluwer, Dordrecht, 2001, pp 59–83. 153
9. V.S. Afraimovich, L. Yu. Glebsky and V.I. Nekorkin, *Stable states in lattice dynamical systems*, Methods of Qualitative Theory and Theory of Bifurcations, University of Nizhny Novgorod, 1991, pp 137–154. 154, 160
10. V. Afraimovich and S.-B. Hsu, *Lectures on Chaotic Dynamical Systems*, AMS/IP Studies in Advanced Mathematics **28**, AMS, Providence, RI, International Press, Somerville MA, 2003. 157, 174
11. L. Yu. Glebsky and A. Morante, *Complete description of weakly coupled chaotic subsystems*, Revista Mexicana de Física **48**, (2002), pp 355–359. 157
12. L. Yu. Glebsky, *Local nonlinear operators and multicontour solutions of homogeneous coupled map lattices*, Mathematical Notes **65** (1999), pp 31–40. 157, 162
13. L.A. Bunimovich and E. Carlen *On the problem of stability in lattice dynamical systems*, J. Diff. Eq. **123**(1995), pp 213–229. 157
14. D. Turzik, M. Dubzova, A. Klic and P. Pokorny, *Application of Bernoulli shift to investigation of stability of spatially periodic solutions in lattice dynamical systems*, Dyn. Syst. **18** (2003), pp 23–33. 157, 165
15. V.S. Afraimovich, L.Yu. Glebsky and V.I. Nekorkin, *Stability of stationary states and spatial chaos in multi-dimentional lattice dynamical systems*, Random and Comp. Dyn. **2** (1994), pp 287–304. 160, 161
16. V.S. Afraimovich and L.A. Bunimovich, *Simplest structures in coupled map lattices and their stability*, Random and Comp. Dyn. **1** (1992-1993), pp 422–444. 160
17. V.S. Afraimovich and V.I. Nekorkin, *Stable states in chain models of unbounded non-equilibrium media*, Mathematicheskoe Modelirovanie **4** (1992), pp 83–95 (in Russian). 160
18. B. Fernandez and A. Morante, *On the stability of periodic orbits in lattice dynamical systems*. Dyn. Syst. **16** (2001), pp 247–252. 160
19. M.I. Rabinovich and M.M. Sushchik, *Regular and chaotic dynamics of structures in liquid flows*, Soviet Phys Uspekhi **33** (1990), pp 1–64 (in Russian) . 160
20. V.S. Afraimovich and S.-N. Chow, *Topological spatial chaos and homoclinic points of $\mathbb{Z}^d$-action in lattice dynamical systems* Jpn. J. Indust. Appl. Math. **12** (1995), pp 1–17. 160, 162

21. S. Zhou, *Attractors for the first order dissipative lattice dynamical systems* Physica D **178** (2003) pp 51–56. 161
22. V.S. Afraimovich, S.N. Chow and W. Shen, *Hyperbolic homoclinic points on $\mathbb{Z}^d$-action in lattice dynamical systems*, Int. J. of Bifurcation and Chaos **6** (1996), pp 1059–1075. 162, 163, 164
23. V.S. Afraimovich and L.A. Bunimovich, *Density of defects and spatial entropy in lattice dynamical systems* Physica D **80** (1995), pp 272–288. 162
24. M. Jiang and R. de la LLave, *Smooth dependence of thermodynamic limits of SRB measures*, Comm. Math. Phys. **211** (2000), pp 303–333. 154, 174
25. R. Horn and U. Johnson, *Matrix Analysis*, Cambridge University Press, 1986. 160
26. V.S. Afraimovich and M.I. Rabinovich, *Complex spatial-temporal dynamics of chain models for flow systems*, in: K. Kaneko, ed., *Theory and application of Coupled Map Lattices* Wiley, Chichester, 1993, pp 117–134. 165
27. V.S. Afraimovich and Ya.B. Pesin, *Hyperbolicity of infinite-dimensional drift systems*, Nonlinearity **3** (1990), 1–19. 165
28. I.S. Aranson, V.S. Afraimovich and M.I. Rabinovich, *Stability of spatially-homogeneous chaotic regimes in unidirectional chains*, Nonlinearity **3** (1990), pp 639–651. 166
29. B. Fernandez, *Global synchronization in coupled map lattices*, preprint CNRS, (2002). 167
30. S. Smale, *Morse inequalities for a dynamical system*, Bull. Amer. Math. Soc. **66** (1960), pp 43–49. 164
31. J. Palis, *On Morse–Smale dynamical systems*, Topology **8** (1968), pp 385–404. 164
32. V.S. Afraimovich and L.P. Shilnikov, *On critical sets of Morse–Smale systems*, Trudy Moscov. Mat. Ob. **28** (1973), pp 181–214 (in Russian). 164
33. I. Aranson, D. Golomb and H. Sompolinsky, *Spatial coherence and temporal chaos in macroscopic systems with asymmetric couplings*, Phys. Rev. Letters **68** (1992), pp 3495–3498. 167
34. V. Afraimovich and Y. Pesin, *Traveling waves in lattice models of multidimensional and multicomponent media: I. General hyperbolic properties*, Nonlinearity **6** (1993), pp 429–455. 169, 170, 173
35. M. Brin and Ya. Pesin, *On Morse-Smale endomorphisms* in: *Sinai Moscow Seminar in Dynamical Systems*, Amer. Math. Soc, Transl. Ser. 2 **171** AMS, Providence, RI, (1996), pp 35–43. 170
36. V.S. Afraimovich and V.I. Nekorkin, *Chaos of traveling waves in a discrete chain of diffusively coupled maps*, Internat. J. Bifur. Chaos Appl. Sci. Engrg. **4** (1994), pp 631–637. 170, 171
37. D.R. Orendovici and Ya.B. Pesin, *Chaos in traveling waves of lattice systems of unbounded media* in: *Numerical Methods for Bifurcation Problems and Large-Scale Dynamical Systems*, (Mineapolis, MN, 1997), Springer, New York, 2000, pp 327–358. 171, 173
38. M. Benedicks and L. Carleson, *The dynamics of the Hénon map*, Ann. Math. **133** (1991), pp 73–169. 171
39. V.I. Arnold. *Problems on singularities and dynamical systems*, in: *Developments in Mathematics: The Moscow School*, V.I. Arnold and M. Monastyrsky (eds), Chapman and Hall, London, Glasgow, New York. 176
40. A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*, Cambridge University Press, London-New York, 1995. 158, 171, 175
41. V. Afraimovich and M. Courbage, *On the abundance of traveling waves in coupled expanding circle maps*, Amer. Math. Soc. Transl, (2) **200** (2000), pp 15–21 171

42. M. Jiang and Ya.B. Pesin, *Equilibrium measures for coupled map lattices: existence, uniqueness and finite-dimensional approximations*, Commun. Math. Phys. **193** (1998), pp 675–711. 173
43. V. Afraimovich and B. Fernandez, *Topological properties of linearly coupled expanding map lattices*, Nonlinearity **13** (2000), pp 973–993. 172, 174
44. R. Coutinho and B. Fernandez, *Extended symbolic dynamics in bistable CML: existence and stability of fronts*, Physica D **108** (1997), pp 249–255. 173
45. R. Coutinho and B. Fernandez, *Fronts and interfaces in bistable extended mappings*, Nonlinearity **11** (1998), pp 1407–1433. 173
46. R. Coutinho and B. Fernandez, *Fronts in extended systems of bistable maps coupling via convolutions*, Nonlinearity **17** (2004), pp 23–47. 173
47. M. Jiang, *Sinai-Ruelle-Bowen measures for lattice dynamical systems*, J. of Stat. Phys. **111** (2003), pp 863–902. 174
48. J.A. Morales, On structural stability of weakly coupled map lattices, PhD thesis, Universidad Autónoma de San Luis Potosí, San Luis Potosí, México, 2003. 174
49. V. Afraimovich, M. Courbage, B. Fernandez and A. Morante, *Directional entropy in lattice dynamical systems* in: *Progress in Nonlinear Science* **1** (L.M. Lerman, L.P. Shilnikov, eds), Nizhny Novgorod, Institute of Applied Physics, 2002, pp 9–30. 174
50. V. Afraimovich, A. Morante and E. Ugalde, *On density of directional entropy in lattice dynamical systems*, Nonlinearity **17** (2004), pp 105–116. 174
51. Y.A. Pesin and Ya.G. Sinai, *Space-time chaos in chains of weakly interacting hyperbolic mappings.* In: *Dynamical systems and Statistical Mechanics*, Moscow, 1991, RI: AMS, Providence, 1991, pp 165–198. 174

Riddled Basins and Coupled Dynamical Systems

P. Ashwin

Department of Mathematical Sciences, Laver Building, University of Exeter, Exeter EX4 4QE, UK
P.Ashwin@ex.ac.uk

1 Introduction

The collective dynamics of groups of coupled dynamical systems is of great interest for understanding spontaneous pattern formation in biological and many other systems; see for example [58]. One can learn a lot about such systems by first studying idealized cases where the systems are perfectly identical; this approach has been very successful in understanding general properties of synchronization as well as particular applications; see for example [77]. In this chapter we consider how this can lead to the appearance of attractors with riddled basins. These basins appear because symmetries of dynamical systems force the presence of invariant submanifolds; the attractors within invariant manifolds may be only weakly attracting transverse to the invariant manifold and this leads to a basin structure that is, roughly speaking, full of holes.

From a theoretical point of view, this behaviour is of interest because it seems strange or pathological but is in some sense common. From a practical point of view, this behaviour points towards the presence of extreme sensitivity of the dynamics to noise, also called 'bubbling' of attractors. Most interestingly, if we consider generic dynamics within a class of symmetric systems, riddled basins can appear as a robust phenomenon; they can be persistent for open sets of parameters of the system.

For the remainder of this section we briefly discuss basins of attraction and a motivating example of a piecewise linear map with an explicitly computable riddled basin attractor. More general properties of riddled sets and basins are discussed in Sect. 2 including their noise sensitivity. This is followed in Sect. 3 by a discussion of the use of symmetries, ergodic measures and Lyapunov exponents tools for identifying riddled basins; we also discuss anisotropic riddling in Sect. 4 along the lines of [7]. Finally in Sect. 5 we outline a few open problems related to riddling phenomena.

P. Ashwin: *Riddled Basins and Coupled Dynamical Systems*, Lect. Notes Phys. **671**, 181–207 (2005)
www.springerlink.com

1.1 Attractors for Smooth Maps on Compact Manifolds

Qualitative dynamics is about ignoring information. Since one cannot find explicit solutions to all but the simplest dynamical systems, the first step is usually to concentrate on what happens "eventually", i.e. for the asymptotic dynamics, and this leads to several possible definitions of attractor. We will discuss a couple of these notions, though there are other notions of attractor based on invariant measures are also very useful; see for example [29, 71]. We refer the reader to the other chapters in this volume for further discussion and examples of attractors in coupled systems as well as [61].

We concentrate on dynamical systems generated by iterated maps

$$f : M \to M$$

where M is a compact manifold and f is a smooth map generating a dynamical system on iteration, though in some cases we may drop assumptions of smoothness or may wish to consider flows F_t with $t \in \mathbb{R}$. There are some comments on generalizations of this to other less restrictive cases in later sections, in particular in Sect. 5.

The (forward) trajectory through $x \in M$ is the set $\{f^n(x) \ : \ n = 0, 1, 2, \cdots\}$ and we say a set $A \subset M$ is (forward) invariant if $f(A) = A$. Define the ω limit set by

$$\omega(x) = \bigcap_{n>0} \overline{\{f^m(x) \ : \ m > n\}} \, .$$

This represents the set of points that the orbit of x accumulates on as $t \to \infty$.

Note that the set $\omega(x)$ is invariant under f. To see this, consider $y \in \omega(x)$; then there is a sequence n_k such that $f^{n_k}(x) \to y$. Continuity of f means that

$$f^{n_k+1}(x) \to f(y)$$

and so $f(y)$ is also in $\omega(x)$. Hence $\omega(x)$ is (forward) invariant.

In cases where f is invertible one can apply the same considerations to $\alpha(x)$, the limits of $f^{-n}(x)$ as $n \to \infty$. An attractor is in some sense the smallest set that contains all $\omega(x)$ limits that one care about; since ω-limit sets are invariant we only really need to consider invariant sets as candidates for attractors. Given an invariant set A, consider the set of points whose orbits are asymptotic to A

$$\mathcal{B}(A) = \{x \in M \ : \ \omega(x) \subset A\}$$

which is the *basin of attraction* of A.

For the definition of attractor given by Milnor [56] we need a Lebesgue equivalent background measure on M that we denote by $\ell(\cdot)$ and by $\ell(A) > 0$ we include the possibility that it is infinite. Recall that Lebesgue measure is simply a generalization of length/area/volume such that one can measure

many sets constructed from infinite unions and intersections of open sets. We say the compact invariant set A is a *weak attractor* if (M1) holds, it is a *Milnor attractor* if (M1) and (M2) hold, and a *minimal Milnor attractor* if (M1) and (M2') hold [16], where:

(M1) $\ell(\mathcal{B}(A)) > 0$.
(M2) For any proper compact invariant subset $A' \subset A$ we have

$$\ell(\mathcal{B}(A) \setminus \mathcal{B}(A')) > 0.$$

(M2') For any proper compact invariant subset $A' \subset A$ we have

$$\ell(\mathcal{B}(A')) = 0.$$

for further comments on these notions see also [24].

One can think of a Milnor attractor as the smallest compact set that attracts all initial conditions except for a set of zero measure with respect to some natural "background measure". As shown in [56] Milnor attractors can be constructed by examining the likely limit sets for positive measure subsets $S \subset M$; these are compact sets $\Lambda(S)$ that are the smallest such that $\omega(x) \subset \Lambda$ except for a zero measure set of $x \in S$.

From the point of view of numerical simulations of a dynamical system, Milnor attractors are just as reasonable a definition of attractor as for example an asymptotically stable attractor; recall that an *asymptotically stable attractor* is an A such that (i) for any open set U containing A there is an open set V containing A such that $x \in V$ implies that $f^n(x) \in U$ for all n and (ii) $\omega(x) \subset A$ for all $x \in V$. An A that satisfies only (i) is Lyapunov stable. Note that an asymptotically stable attractor must be a weak Milnor attractor with a basin than contains an open set.

1.2 A Motivating Example

Simple examples of systems with riddled basin attractors can be found by considering skew product dynamical systems. A *direct product* of two maps g and h is simply the map obtained by $f(x, y) = (g(x), h(y))$; g and h are referred to as factors; projecting onto one of the coordinates gives a map that is well-defined. A *skew product* is a map of the form

$$(x, y) \mapsto f(x, y) = (g(x), h(x, y)) \tag{1}$$

that has only one factor, in this case g. In cases where $h(x, y) = y\hat{h}(x, y)$ the map (1) has an invariant subspace $N = \{(x, 0)\}$. We consider as in [6] the piecewise linear map on $(x, y) \in [0, 1] \times [0, \infty)$ of the form (1) where

$$g(x) = \begin{cases} \alpha^{-1}x & \text{for } 0 \le x < \alpha \\ (1-\alpha)^{-1}(x-\alpha) & \text{for } \alpha \le x < 1 \end{cases} \tag{2}$$

$$h(x, y) = \begin{cases} \gamma y & \text{for } y < 1 \text{ and } 0 \le x < \alpha \\ \gamma^{-1} y & \text{for } y < 1 \text{ and } \alpha \le x \le 1 \\ 1 + \beta(1-y) & \text{for } y \ge 1\ . \end{cases} \tag{3}$$

We will assume that $\gamma > 1$, $0 < \alpha < \frac{1}{2}$ and $-1 < \beta < 0$ are fixed. Varying α through $1/2$ allows one to observe a blowout bifurcation of an attractor $A = N$. Varying β allows one to change the dynamics between supercritical and subcritical scenarios; for more details of the other parameter values, see [6]. In the case $-1 < \beta < 0$, any trajectory that arrives in $y \leq 1$ will stay there for ever.

First, observe that this map has an invariant subspace $y = 0$ on which the dynamics is chaotic on this set A in the strongest sense that is commonly in use; the dynamics has Lebesgue measure as an ergodic invariant measure for which the dynamics is Bernoulli.

Now pick any $x \in [0, 1]$ and consider its itinerary under the "skewed doubling map" g. This defines a symbol sequence $\{s_i\}_{i=0,1,2\ldots}$ where $s_i = 0$ if $0 \leq f^i(x) < \alpha$ and $s_i = 1$ if $\alpha \leq f^i(x) \leq 1$. Define

$$l_k = \#\{0 \leq j < k \ : \ s_j = 0\}, \ r_k = \#\{0 \leq j < k \ : \ s_j = 1\},$$

the number of times that the itinerary of x is resp. to the left/right of $x = \alpha$. For almost all x we can use the fact that Lebesgue measure on $[0, 1]$ is invariant and ergodic under g to conclude that the following sequences converge

$$\lim_{k\to\infty} \frac{l_k}{k} = \alpha, \ \lim_{k\to\infty} \frac{r_k}{k} = 1 - \alpha \tag{4}$$

for almost all $x \in [0, 1]$. If we now define

$$M_k(x) = \gamma^{l_k - r_k}$$

then as long as $\alpha < \frac{1}{2}$ we have by (4) that $\lim_{k\to\infty} \frac{1}{k}(l_k - r_k) = 2\alpha - 1 < 0$ so that $\lim_{k\to\infty} l_k - r_k = -\infty$. Hence

$$\lim_{k\to\infty} M_k(x) = \lim_{k\to\infty} \exp\left[(l_k - r_k)\ln\gamma\right] = 0$$

for almost all x. By comparing with (3) one can verify that

$$f^k(x, y) = (g^k(x), M_k(x)y) \tag{5}$$

as long as $M_k(x)y$ does not exceed 1. We define

$$Y(x) = \max\left(1, (\sup_{k\geq 0} M_k(x))^{-1}\right)$$

For almost all x we have $0 < Y(x) \leq 1$. If $y < Y(x)$ then (5) holds for all $k \geq 0$ because $M_k(x)y$ will never exceed 1.

In summary, there is a function $Y(x)$ with $Y(x) > 0$ for almost all x that describes the basin of attraction of A:

$$\mathcal{B}(A) = \{(x, y) \ : \ 0 \leq y < Y(x)\}.$$

Clearly, the Lebesgue measure of $\mathcal{B}(A)$ must be greater than zero; it is simply $\ell(\mathcal{B}(A)) = \int_{[0,1]} Y(x)\,dx$, and A is a minimal Milnor attractor because almost all x have g orbits that are dense in $[0,1]$.[1] However the function $Y(x)$ is highly non-smooth and it is this that makes the basin $\mathcal{B}(A)$ in fact riddled as shown in Fig. 1.

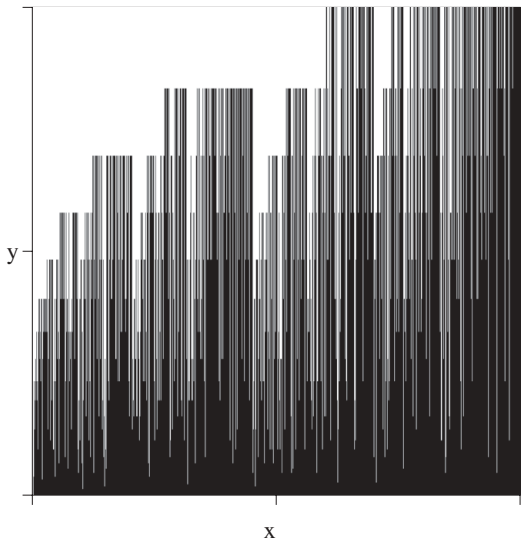


Fig. 1. The *black* set shows a numerical approximation of the riddled basin for the attractor in $(x, 0)$ with $\alpha = 0.45$, $\gamma = 1.2$ and $\beta = -1$ for the map (1,2,3). The *box* shows points $(x, y) \in [0,1]^2$ while initial conditions in the white set are ejected to $y \geq 1$

For this example we can compute the measure within the basin as in [6]. Let $\epsilon = 2\alpha - 1$ and note that $\alpha < 1/2$ implies that $\epsilon < 0$ in what follows. We partition $[0,1]^2$ into a set of strips

$$I_n = [0,1] \times (\gamma^{-n-1}, \gamma^{-n}),\ \ n = 0, 1, 2, \ldots$$

where the strip I_n has height $\gamma^{-n-1}(\gamma - 1)$. The form of the map means that it is conjugate to a mapping on the I_n defined by

$$T(x, n) = (g(x), m(x, n))$$

where

$$m(n, x) = \begin{cases} n - 1 & \text{if } 0 \leq x < \alpha \text{ and } n > 0 \\ n + 1 & \text{if } \alpha \leq x \leq 1 \text{ and } n > 0 \\ 0 & \text{if } n = 0 \end{cases}$$

[1] Recall that a sequence of points x_n is *dense* in a metric space M if any open set in M contains a point in the sequence.

which can be viewed as a biased random walk on $\mathbb{N} \cup \{0\}$ with 0 being an absorbing state. Let q_n, $n > 0$, be the probability of arriving at the state 0; then

$$q_n = \frac{1}{2}(1+\epsilon)q_{n-1} + \frac{1}{2}(1-\epsilon)q_{n+1}$$

which has solution $q_n = \Lambda^{-n}$ where $\Lambda = (1-\epsilon)(1+\epsilon)$ will satisfy $\Lambda > 1$. Hence the measure of points that are attracted to the invariant subspace is given by

$$\ell(\mathcal{B}(A)) = \sum_{n=1}^{\infty} \gamma^{-n-1}(\gamma-1)(1-\Lambda^{-n}) = \frac{\gamma(\Lambda-1)}{\Lambda\gamma-1} \cdot$$

Clearly $\ell(\mathcal{B}(A)) \to 0$ as $\epsilon \to 0^-$. We illustrate in Fig. 2 the structure of the complement of the basin of attraction; note that there are "tongues" of instability that come down to touch the x-axis at all points $(x, 0)$ such that the itinerary of x ends in an infinite number of 0s. Since this set is dense, we can conclude that the basin of attraction of $y \geq 1$ is dense and all points must exit after a finite time, it follows from this that this set is open and dense. In terms of $Y(x)$ this means that $Y(x) = 0$ on a dense set in $[0, 1]$; since $Y(x)$ is almost everywhere positive, $Y(x)$ is discontinuous on a dense set on $[0, 1]$. In fact, $Y(x)$ is upper semicontinuous at almost all x and the set $\{(x, y) \ : \ y > Y(x)\}$ is open and dense in $\mathbb{R}^2$.

One might suspect that this behaviour is caused by the presence of discontinuities in the map (1) but in fact this is not the case; similar basins appear

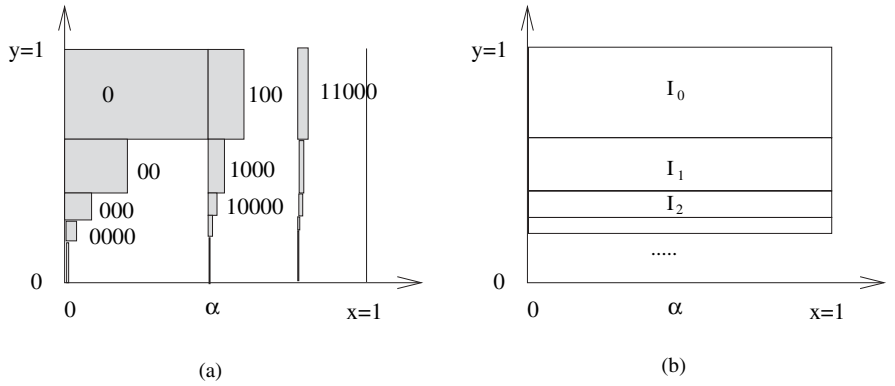


Fig. 2. (**a**) The structure of the basin of attraction of A for the map (1,2,3) discussed in the text. For $\alpha < 0$ the invariant set A in $y = 0$ has a basin with positive measure whose complement is open and dense and whose structure is shown by the *shaded* set. The coding indicates the x-itinerary of those points in $y < 1$ before they are expelled to $y \geq 1$. By including all possible finite words that occur before expulsion one obtains a set that is dense and open in $[0, 1]^2$ yet which does not have full measure. (**b**) shows the strips I_n for this map (see text)

naturally in smooth and even invertible maps and seem to be robust in many systems with symmetries or invariant subspaces.

1.3 Related Notions of Basin Complexity

A somewhat simpler concept than riddled basin, is that of a *fractal basin boundary*. It has been observed since the work of Julia and Fatou for complex maps, that even linearly stable equilibria e may have basins of attraction $B = \mathcal{B}(\{e\})$ with boundary such that the Hausdorff dimension $\dim_H(\partial B)$ is not an integer; for example, see [32, chapter 14] (recall that $\partial B = \overline{B} \setminus \text{Interior}(B)$). One should stress however that for riddled basins, not just the boundary has "fractal" properties, but the whole set is inseparable from its boundary and in fact $\partial B = B$ up to a set of zero measure.

Another notion of attraction that is weaker than asymptotic stability but is stronger than Milnor attraction is that of *essential asymptotic stability* [54]; we say an invariant set A is e.a.s. if it is asymptotically stable if one excludes a set of small measure compared to small neighbourhoods of A; more precisely if there is a set $\mathcal{S}$ such that for any neighbourhood U of A and any $0 < a < 1$ there is a neighbourhood V of A with $\ell(V \setminus S)/\ell(V) > a$ where $x \in V \setminus \mathcal{S}$ implies that $f^n(x)$ remains in U and is asymptotic to A. Such attractors are found quite commonly and robustly in heteroclinic networks that have lost asymptotic stability. However in this case the basin of attraction of A may still be an open unriddled set.

2 Riddled Sets and Riddled Basins

In contrast to attractors with fractal basin boundaries, a riddled basin is "fractal" everywhere. However, since a basin of attraction must have positive measure within phase space, in fact it must have Hausdorff dimension equal to that of phase space and so cannot have non-integer dimension. It is a "fat fractal" in the terminology of [30, 34] in that it contains a dense set of holes. In the following we will use the definitions as in [16] though we note that there are several possible equivalent definitions. We denote by $B_\delta(x) = \{y \in M \ : \ |y - x| < \delta\}$ the open δ-ball about x in M.

A *riddled* subset $A \subset \mathbb{R}^n$ is a measurable set with the property that for any $\delta > 0$ and any $x \in A$ we have

$$\ell(A \cap B_\delta(x)) > 0 \quad \text{and } \ell(A^c \cap B_\delta(x)) > 0 \, . \tag{6}$$

More generally given any set A with positive measure one can define its *riddled component*

$$A_{rid} = \{x \in A \ : \ \ell(B_\delta(x) \cap A)\ell(B_\delta(x) \cap A^c) > 0 \quad \text{for all } \delta > 0\} \, .$$

We use this to distinguish cases of full riddling ($\ell(A \setminus A_{rid}) = 0$), lack of riddling ($\ell(A_{rid}) = 0$), with partial riddling being any other case. If A is the closure of an open set with smooth boundary then $\partial A = A_{rid}$.

Sometimes one would like to discuss a similar condition to (6) but rather than comparing A and its complement A^c one would like to compare to disjoint subsets A and B of $\mathbb{R}^n$. If A and B are disjoint and both have positive measure then we say A is *riddled with* B in the case that almost all $x \in A$ and all $\delta > 0$ have

$$\ell(B_\delta(x) \cap A)\ell(B_\delta(x) \cap B) > 0 \, .$$

In the case that A is riddled with B and B is riddled with A we say they are *intermingled*.[2]

One can show that the riddled component of an invariant set A is invariant for a large class of maps. We say f is of type (P) if it is continuous, a local homeomorphism and nonsingular (i.e. for any V, $\ell(V) = 0$ if and only if $\ell(f(V)) = 0$). The following result taken from [16] can be generalised to cases where f is almost everywhere a local homeomorphism or to some classes where f is invertible but discontinuous on a set of zero measure [12].

Theorem 2.1. *Suppose that $f : M \to M$ is of type (P) and V is invariant, then V_{rid} is invariant.*

Proof. Consider $x \in V$ and choose a neighbourhood U_1 of x and U_2 of $f(x)$ such that $f : U_1 \to U_2$ is a homeomorphism. Consider any $\delta > 0$ such that $B_\delta(x) \subset U_1$ and $B_\delta(f(x)) \subset U_2$. Continuity of f means that we can find $\delta > \epsilon > 0$ such that $f(B_\epsilon(x)) \subset B_\delta(f(x))$. Hence $\ell(B_\delta(f(x)) \cap V^c) \geq \ell(f(B_\epsilon(x)) \cap V^c)$; by considering the local inverse we see that there is a $\delta > 0$ such that $\ell(B_\delta(x) \cap V^c) = 0$ if and only if there is an $\epsilon > 0$ such that $\ell(B_\epsilon(f(x)) \cap V^c) = 0$. Applying the same argument to V in place of V^c means that x is in V_{rid} if and only if $f(x)$ is. □

This result can adapted to the case where f is almost everywhere a local homeomorphism in which case we conclude that V_{rid} is invariant up to a set of zero measure.

2.1 Characterization of Riddled Basins

One approach to characterizing dynamical invariants of riddled basins has been to compute their *uncertainly exponent*; [34, 62]. This allows one to compute a quantity that characterizes the "riddledness" of the basin or more generally of a "fat fractal". We define this as in [62] for a riddled subset of $\mathbb{R}^2$.

[2] If two positive measure sets A and B are intermingled then they cannot be evenly distributed in the following sense; the Lebesgue density theorem [82, p107] or [32, p69] implies that almost all points in A are points of density for A in the sense that for almost all $x \in A$, $\lim_{\epsilon \to 0} \ell(B_\epsilon(x) \cap A)/\ell(B_\epsilon(x)) = 1$. Hence almost all points in $A \cup B$ are density points for only one of the sets A or B.

Pick a typical line transverse to the basin and choose two points within 2ϵ of each other. If we can estimate the probability p that one point is in the basin and the other is not by $p \sim \epsilon^{\phi}$ we say that there is an uncertainty exponent ϕ; in [62] it is shown that one can estimate this as the ratio $\phi = (\lambda^{\perp})^2/4D\lambda^{||}$ where $\lambda^{\perp}$ and $\lambda^{||}$ are the transverse and tangential Lyapunov exponents and D is the rate of convergence of variance of finite time Lyapunov exponents.

Another (simpler) characterization of a riddled basin proposed by [62] is the scaling of the measure near the attractor; this corresponds to taking a line at a distance y from the attractor; in many cases the measure of the basin intersected with this line scales as y^{η}. This is estimated in [62] as being $\eta \sim |\lambda^{\perp}|/D$.

An interesting observation is that the Kaplan-Yorke formula relating dimension of attractor to Lyapunov dimension [44] can fail for riddled basin attractors; one characteristic of blowout bifurcations is that they are associated with a sudden regain of the validity of that formula.

2.2 Locally and Globally Riddled Basins

There are cases where an attractor has a basin with full measure in some open neighbourhood but still a form of riddling in the convergence towards the attractor. More precisely, consider any open neighbourhood U of A and define the *basin of A relative to U* to be

$$\mathcal{B}_U(A) = \{x \in M \ : \ \omega(x) \subset A \quad \text{and } f^n(x) \subset U \quad \text{for all } n \geq 0\} \ .$$

We say the basin of A is *locally riddled* if there is a neighbourhood U of A such that $\mathcal{B}_U(A)$ is riddled [16] (this is a stronger assumption than that given in [9] where it is only assumed that the riddled component of $\mathcal{B}_U(A)$ is dense in A). It is *globally riddled* in the case that U can be chosen to be equal to M.

2.3 Riddling and Noise Sensitivity; Bubbling of Attractors

A useful model for noise in iterated maps consists of adding an independent uniformly distributed random variables to all components at every iteration. For attractors with riddled basins, this can give rise to discontinuous behaviour in the support of attractors as the noise goes to zero and was called *bubbling* in [8].

More precisely, suppose that $A \subset N \subset M$ with A a Milnor attractor for $f : M \to M$ and the basin of A is locally riddled (say $\mathcal{B}_U(A)$ is riddled with U compact in M) but which is asymptotically stable in the invariant subspace N. Consider the perturbed map

$$x_{n+1} = f_{\sigma}(x) = f(x_n) + \sigma\xi_n \tag{7}$$

where ξ_n is a vector of i.i.d. random variables uniformly distributed in $[-0.5, 0.5]$. Suppose that f_{σ} has an attractor A_{σ} in the sense that A_{σ} is the

smallest compact set that contains all limit points for realizations of the noise ξ_n and all x in some open region.

We say the attractor A for f is *stable to noise* if $A_\sigma \to A$ in the Hausdorff metric as $\sigma \to 0$. This is a weaker notion than for example stochastic stability discussed in [18]; the latter considers convergence of measures for the perturbed system to a natural measure for the noise-free system. Define the *unstable set* of A to be

$$\mathcal{U}(A) = \{x \ : \ \exists \{x_{-n}\} \quad \text{with } f(x_{n-1}) = x_{-n} \ \& \ \lim_{n\to\infty} d(x_{-n}, A) = 0\}$$

i.e. the set of points that have a backward trajectory from A (this definition is simpler if f is invertible); clearly $\mathcal{U}(A) \supset A$. If A is stable to noise then $\mathcal{U}(A) \subset A$. Riddled basin attractors typically have $\mathcal{U}(A) \not\subset A$ and so are not stable to noise.

We will explore further properties of bubbling (in the presence of several invariant subspaces) in Sect. 4.

3 Symmetry, Transverse Stability and Riddling

Consider a smooth iterated mapping $f : M \to M$ with $M = \mathbb{R}^m$ some Euclidean space and write $\ell_N(.)$ to denote Lebesgue measure on N a linear subspace of M. Suppose that f commutes with (is equivariant for) the action of some finite matrix group Γ acting on M (we only consider finite groups here though similar results can be obtained for compact groups). This means that for any $x \in M$ and for any $g \in \Gamma$ we have that

$$g.f(x) = f(g.x) \ .$$

In general this will force a number of linear subspaces of M to be f-invariant. More precisely, given any subgroup Σ of Γ we define the *fixed point subspace*

$$\mathrm{Fix}(\Sigma) = \{x \in M \ : gx = x \quad \text{for all } g \in \Sigma\}$$

which is f-invariant. Not all subgroups give rise to distinct fixed point subspaces; those that do are the *isotropy subgroups* of the action of Γ on M; these are the subgroups $\Sigma(x)$ with $x \in M$ such that $\Sigma(x) = \{g \in \Gamma \ : \ gx = x\}$; these are precisely the possible symmetries of points in M. We therefore obtain, for a given group action, a finite number of linear subspaces $N_i \subset M$, $i = 1 \cdots n$ and isotropy subgroups $\Sigma_i \leq \Gamma$ such that $N_i = \mathrm{Fix}(\Sigma_i)$ and so $f(N_i) \subset N_i$. Note that $N_i \subset N_j$ if and only if $\Sigma_i \geq \Sigma_j$. Observe also that $N_i \cup N_j = N_k$ for some k is also invariant. Conversely, if Σ_k is the smallest isotropy subgroup that contains both Σ_i and Σ_j then $N_i \cup N_j = N_k$.

Two subgroups Σ_1, Σ_2 are *conjugate* if there is a $g \in \Gamma$ such that $\Sigma_1 = g^{-1}\Sigma_2 g$. The next basic result states that conjugate subgroups have fixed point spaces that are mapped onto each other by the group. More precisely we have the following elementary result.

Lemma 3.1. *Suppose that Σ is an isotropy subgroup and $g \in \Gamma$. Then $g^{-1}\Sigma g$ is an isotropy subgroup and* $\mathrm{Fix}(g^{-1}\Sigma g) = g^{-1}\mathrm{Fix}(\Sigma)$.

Proof. Note that

$$\begin{aligned}\mathrm{Fix}(g^{-1}\Sigma g) &= \{x \in M \ : \ g^{-1}hgx = x \quad \text{for all } h \in \Sigma\} \\ &= g^{-1}\{gx \in M \ : \ hgx = gx \quad \text{for all } h \in \Sigma\} = g^{-1}\mathrm{Fix}(\Sigma)\ .\end{aligned}$$

□

There has been some work on symmetries of general chaotic attractors in symmetric systems; see [33] for an overview, and [17, 19, 25, 27, 55]. Other phenomena that can appear include intermittent dynamics between states with a variety of different symmetries; for example [13, 14, 31].

3.1 Example; Invariant Subspaces for Four Globally Coupled Maps

Consider the map on $\mathbb{R}^4$ defined by

$$x_i' = (1-\epsilon)f(x_i) + \frac{\epsilon}{4}\sum_{j=1}^{4} f(x_j) + \sigma_i\eta_i \tag{8}$$

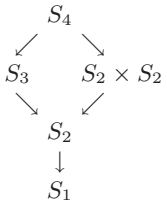
for $i = 1, \ldots, 4$, where at each time-step each η_j is an independent random variable that is uniformly distributed on $[-0.5, 0.5]$. The local map $f(x) = 1 - ax^2$ is a quadratic map and ϵ is the coupling strength. The noise in the ith component can be controlled by setting σ_i nonzero.

The system (8) has been studied by several authors including notably Kaneko [42] although only with isotropic noise perturbations, and displays a wide range of synchronization and chaotic behaviour. Taborev et al. [79] have also recently looked at the noise-free case of $n = 3$ cells in some detail.

This map has the symmetry of all permutations on n objects; it is equivariant under the action of the group S_4 given by permutation matrices on $\mathbb{R}^4$. There are many invariant subspaces corresponding to isotropy subgroups; these can be characterized by partitions of $\{1..4\}$ into $1 \le m \le 4$ groups of identical cells and after identifying conjugate subgroups the possible isotropy subgroups are conjugate to one of the *partial clustering states* S_1, S_2, $S_2 \times S_2$, S_3 and S_4. These have fixed point subspaces that are given by coordinates within one partition being equal; for example if H is the subgroup generated by the two-cycles (12) and (34) then $\mathrm{Fix}(H) = \{(x, x, y, y)\}$.

One can order the possible subspaces by $\Sigma_1 < \Sigma_2$ if and only if $\mathrm{Fix}(\Sigma_2) \subset \mathrm{Fix}(\Sigma_1)$; for the case $\Gamma = S_4$ this gives the containments as shown in the *isotropy lattice* in Table 1. Note that for n larger the isotropy lattice of S_n becomes much more complicated.

Table 1. The lattice of isotropy subgroups up to conjugacy for the symmetry group S_4 of four globally coupled maps. The arrows indicate containment of subgroups; the isotropy subgroups can be interpreted as cluster states for the coupled maps



3.2 Symmetries and Lyapunov Exponents

Now suppose we have a compact invariant set $A \subset N \subset M$ with N a linear invariant subspace. Assume that A supports an ergodic f-invariant probability measure μ_{nat} that is a *natural measure*, i.e. there is a positive measure set (with respect to Lebesgue measure on N) such that points in these sets have ergodic averages determined by μ; i.e. such that

$$\lim_{k\to\infty} \frac{1}{k} \sum_{j=0}^{k-1} \phi(f^j(x)) = \int_A \phi(y)\, d\mu(y)$$

for any continuous $\phi : N \to \mathbb{R}$ and a positive measure set of $x \in N$. We also define

$$\mathrm{Erg}(A) = \{\mu \ : \quad \text{ergodic measures with support on } A\}.$$

With respect to any ergodic measure $\mu \in \mathrm{Erg}(A)$ we define the *Lyapunov exponents* (L.E.s)

$$\lambda(x, v) = \lim_{n\to\infty} \frac{1}{n} \log \frac{|Df^n(x)v|}{|v|} \ .$$

Oseledec's theorem [29, 53, 60, 66, 71] implies that for μ-almost all x and any v this limit converges and may take one of only a finite number of possible values $\lambda_i(\mu)$, $i = 1, \ldots, m$ (we count multiplicity by the dimension of the subspace of v that give this value for typical x). Recall that given any ergodic measure with support contained within an invariant subspace N, the Lyapunov exponents (L.E.s) will split into two groups [9, 23]; the *tangential L.E.s* $\lambda_j^{||}(\mu)$ and the *transverse L.E.s* $\lambda_j^{\perp}(\mu)$; the former correspond to perturbations v within N (cf the Sacker-Sell spectrum [75]). We assume that for any ergodic μ the transverse L.E.s are ordered greatest first; $\lambda_1^{\perp}(\mu) \geq \lambda_2^{\perp}$ etc. Define

$$\Lambda_{\max} = \sup_{\mu\in\mathrm{Erg}(A)} \lambda_1^{\perp}(\mu), \ \Lambda_{nat} = \lambda_1^{\perp}(\mu_{nat}).$$

In the case that $\Lambda_{\max} > 0$ and $\Lambda_{nat} < 0$ we will obtain an attractor A whose basin is locally riddled and possibly globally riddled, and at least there will be

a neighbourhood U for which the attractor have a relative basin $\mathcal{B}_U(A)$ whose riddled component includes A. It is however more difficult to come up with necessary conditions for globally riddled basins that are not very restrictive. See [7, 9, 16, 24] for some conditions that imply local riddling.

We say a Milnor attractor is *regular* if for any open neighbourhood U of A then the measure of the basin of A relative to U is positive; $\ell_M(\mathcal{B}_U(A)) > 0$. A result in [1] implies that any Milnor attractor that is a uniformly hyperbolic within N with $\Lambda_{nat} < 0$ will be regular.

Note that for uniformly hyperbolic A (within N) all ergodic measures on A are limit points of sequences of periodic measures supported within A [78] and so for such sets the existence of a measure with positive transverse Lyapunov exponent implies the existence of periodic points with L.E.s that are arbitrarily close to that of the natural measure.

3.3 Examples of Riddled and Intermingled Basins

Since the phenomena of riddled basins was uncovered [1], riddled basins have been found in a range of applications, for example learning dynamical systems [59], coupled chaotic oscillators [8], mechanical systems [83] and electronic systems [38] and especially coupled maps, e.g. [36] where the literature is too extensive to list; see other chapters in this book.

As an example, we consider here a smooth map from [11] with a riddled basin attractor:

$$f(x, y) = (rx(1 - x) + sxy^2, \nu e^{-x} y + y^3) \tag{9}$$

with $r = 4$, $s = 0.3$ and $\nu = 1.5$. This has an attractor within $y = 0$ given by $A = [0, 1] \times \{0\}$ on which the dynamics is that of a logistic map with $r = 4$.[3] Evaluating the transverse Lyapunov exponent with respect to any ergodic measure μ within A amounts to computing the integral

$$\lambda^{\perp}(\mu) = \int (-x + \ln(\nu))\, d\mu(x) = \ln \nu - \int x\, d\mu(x).$$

Observe that (9) has a fixed point in A at $(0, 0)$ has the largest possible $\lambda^{\perp} = 0.40546511$ meaning that $\Lambda_{\max} = 0.40546511$, while the natural measure gives $\Lambda_{nat} = -0.094534$. We can use the topological conjugacy of the map f on A to a doubling map to conclude that there is a dense set of preimages of the fixed point at $(0, 0)$; hence we expect the basin of attraction of A to be locally riddled. In this case we can observe that it is globally riddled; see Fig. 3. Note that unlike the example (1) this map is neither discontinuous nor a skew product, as long as $s \neq 0$.

Examples of intermingled basins have been found in maps [2] (three intermingled basins) and [41] as well as coupled ODEs [28]. We present an example

[3] In this case the basin of A within $y = 0$ contains no points outside of A.

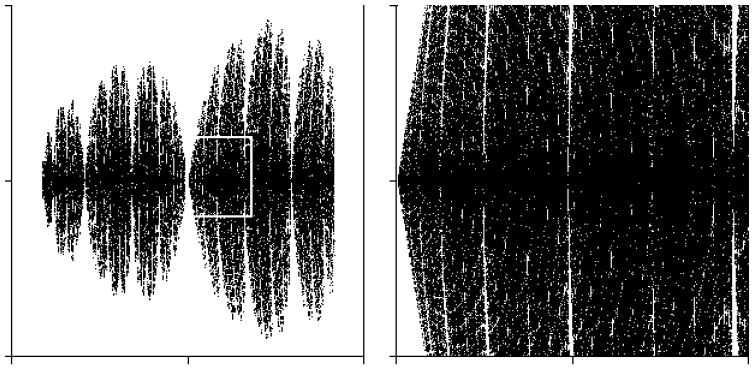


Fig. 3. Numerical approximation of the riddled basin of attraction of an attractor in $y = 0$ for the smooth map (9) with $r = 4$, $s = 0.3$ and $\nu = 1.5$. The basin is shown in black, while the points in white converge to attractors at infinity. The image (**a**) shows the basin in the area $[-0.1, 1.1] \times [-0.6, 0.6]$ while (**b**) shows a zoom into $[0.5, 0.7] \times [-0.1, 0.1]$

in a map due to Ding and Yang [28] of two intermingled basins for a map on $(x, y) \in [-1, 1]^2$ given by $f(x, y) = (x', y')$ where

$$\begin{aligned} x' &= g(x) + \epsilon(g(y) - g(x) + g(y)^3 - g(x)^3) \\ y' &= g(y) + \epsilon(g(x) - g(y) + g(x)^3 - g(y)^3) \end{aligned} \tag{10}$$

for $g(x) = 3.4x(1 - x^2)e^{-x^2}$ and $\epsilon = 0.48$. This map has attractors $A^{\pm}$ in the diagonal $x = y$ on either side of the origin. Figure 4 illustrates the basin of the attractor A^+ in black; the approximation is obtained by computing

$$A_n = f^{-n}([0, 1] \times [-1, 1]).$$

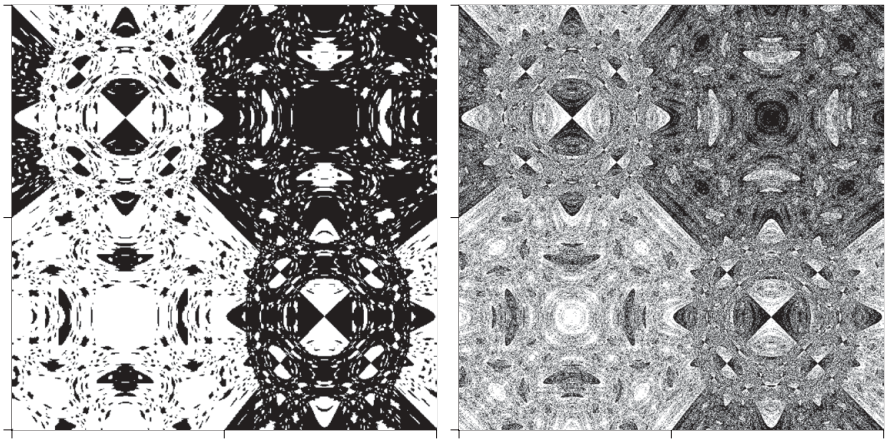
If we assume that $A^{\pm}$ are the only Milnor attractors for f then

$$\bigcap_{m>n} \left(\bigcup_{p>m} A_p \right)$$

will converge (apart from a set of zero measure) to $\mathcal{B}(A^+)$ as $n \to \infty$. The sets A_6 and A_{25} are shown; the Lebesgue density theorem implies that almost all points are points of density for either the white or the black set (though there are clearly do exist points such as $(0, 0)$ that are points of density for neither white nor black set; it is even possible that these are dense in $[-1, 1]^2$).

3.4 Normal and Non-normal Parameters

Most analytical studies of blowout bifurcation rely on being able to vary the dynamics while preserving the dynamics on some invariant subspace. This



(a) (b)

Fig. 4. Numerical approximation of intermingled basins of attraction of an attractor A^+ in $x = y$, $x > 0$ for the map (10) with $\epsilon = 0.48$. The basin is shown in black, while the points in white are an approximation of the set basin of A^-. Both images show the area $(x, y) \in [-1, 1]^2$; (**a**) shows an approximation using 6 iterates while (**b**) uses 25 iterates

allows one to effectively vary the transverse Lyapunov exponents for fixed invariant measures. More precisely, if $f_r : M \to M$ is parameterized smoothly by $r \in \mathbb{R}$ and if $f_r(N) = N$ for all r and some fixed submanifold then we say r is a *normal parameter* if $f_r|_N$ is independent of r. Otherwise we refer to r as a *non-normal parameter*. For normal parameters the Lyapunov exponents will vary with the parameter in a way that may or may not be analytic; see for example [9, 73, 74]. In cases that the normal Lyapunov exponent does vary continuously, the transition to riddling of a basin can be determined by finding the point at which it loses asymptotic stability.

As discussed in [11, 26] the transitions obtained on varying a non-normal parameter will typically be much more complicated than on varying a normal parameter, unless the attractor within N is robust [20]. We refer to the other chapters in this volume for discussion of the blowout bifurcation.

4 Anisotropic Riddling in Coupled System

Following [7] we consider properties of riddling that can appear in the presence of more than one invariant subspace. Riddling of the basin may occur in some directions but not others; we say that riddling is typically *anisotropic* in transverse directions unless they are symmetrically related in ways that we characterize in Theorems 4.1, 4.2. In more general cases the dynamics can display anisotropic sensitivity to noise. We also illustrate internal riddling transitions where the number of directions in which riddling occurs changes.

For nested invariant subspaces N and P with $A \subset N \subset P \subset M$ we define

$$\{\lambda_i^{N,P}(\mu) \ : \ i = 1, \ldots, \dim P - \dim N\}$$

to be the set of possible $\lambda^{\perp}(x, v)$ attained for $v \in P$ and x is typical with respect to the measure μ. We also define

$$\Lambda_{\max}^{N,P} = \sup_{\mu \in \mathrm{Erg}(A)} \lambda_1^{N,P}(\mu)$$

and

$$\Lambda_{nat}^{N,P} = \lambda_1^{N,P}(\mu_{nat}).$$

In the case of symmetries we often obtain attractor within an invariant subspace N that is nontrivially contained within several distinct invariant subspaces P_1, P_2 etc. The L.E.s in different directions P_i can however in certain circumstances be related. In what follows T is some isotropy subgroup of Γ:

Theorem 4.1. *Suppose that $N = \mathrm{Fix}(T)$ and $N \subset P_k = \mathrm{Fix}(\Sigma_k)$, $k = 1, 2$ where $\Sigma_k \subset T$ are isotropy subgroups that are conjugate within T. Suppose that $A \subset N$ is an attractor. Then λ_i^{N,P_k}, $\Lambda_{\max}^{N,P_k}$ and Λ_{nat}^{N,P_k} are independent of k.*

Proof. We write $\Sigma = \Sigma_1$ so that $P_1 = \mathrm{Fix}(\Sigma)$. There is a $g \in T$ such that $P_2 = \mathrm{Fix}(g^{-1}\Sigma g) = g^{-1}\mathrm{Fix}(\Sigma) = g^{-1}P_1$. Now $gx = x$ and G has orthogonal action, so $|gy| = |y|$ for any y. Equivariance of f implies equivariance of the derivative $(Df(gx)gv = gDf(x)v)$. Hence for any $x \in N$ and $v \in P_1$ we have

$$\lambda(g^{-1}x, g^{-1}v) = \lim_{n\to\infty} \frac{1}{n} \log \frac{|Df^n(gx)gv|}{|gv|}$$

$$= \lim_{n\to\infty} \frac{1}{n} \log \frac{|gDf^n(x)v|}{|gv|} = \lambda(x, v).$$

This means that any L.E. λ_i^{N,P_1} is also a L.E. λ_i^{N,P_2} and vice versa. □

The previous result requires that the conjugating element g is in T. More generally one can require that g maps $\mathrm{Fix}(T)$ to itself. This implies that $g \in \mathrm{Norm}(T)$ where $\mathrm{Norm}(T) = \{h \in \Gamma \ : \ hT = Th\}$ is the normalizer of T. In this case the result above can be adapted as long as the measure μ_{nat} has symmetry g (a *symmetry on average* [19]). More precisely,

Theorem 4.2. *Suppose that $N = \mathrm{Fix}(T)$ and $N \subset P_k = \mathrm{Fix}(\Sigma_k)$ for $k = 1, \ldots, l$ where $\Sigma_k \subset T$ are related by $\Sigma_k = g_k^{-1}\Sigma g_k$ for some $g_k \in \mathrm{Norm}(T)$. Suppose that $A \subset N$ is an attractor with natural measure μ_{nat} invariant under action of all g_k. Then λ_i^{N,P_k}, $\Lambda_{\max}^{N,P_k}$ and Λ_{nat}^{N,P_k} are independent of k.*

Proof. This follows as for the previous result on noting that there is a g_k-invariant set of x with full μ_{nat}-measure that has the same L.E.s. at each point. □

4.1 Directions of Riddling

Recall that a basin of attraction $\mathcal{B}(A)$ is *riddled* (in M) [1] if for any $x \in A$ and $\delta > 0$ we have

$$\ell_M(\mathcal{B}(A) \cap B_\delta(x))\ell_M(\mathcal{B}(A)^c \cap B_\delta(x)) > 0$$

where $\ell_M(\cdot)$ denotes Lebesgue measure on M. Similarly, we say the basin of attraction of $A \subset N_0$ is *riddled in the direction* N_i for any invariant subspace N_i where $N_0 \subset N_i$ if

$$\ell_{N_i}(\mathcal{B}(A) \cap B_\delta(x))\ell_{N_i}(\mathcal{B}(A)^c \cap B_\delta(x)) > 0.$$

Observe that it is necessary for $\mathcal{B}(A) \cap N_i$ to have positive ℓ_{N_i}-measure in order to get riddling.

Suppose that A is a Milnor attractor such that $A \subset N_0 \subset N_1 \subset M$. If A is riddled in M then it need not be riddled in either N_i. In fact examples discussed in [1, 9] have attractors that are riddled in M but asymptotically stable and therefore unriddled in the largest linear subspace N that contains the attractor. What we emphasise here is that it may be unriddled in a *larger* invariant subspace. The same holds even if A is not an attractor but a chaotic saddle in M.

Suppose a Milnor attractor A in M has $\Lambda_{nat} < 0 < \Lambda_{\max}$ for the system restricted to an invariant subspace N' with $N \subset N' \subset M$. Then the same inequality holds for the full system and will imply riddling in the full basin, as long as A is an attractor in M.

One can find systems $f : M \to M$ with an invariant set A contained in an invariant subspace N such that A is an attractor in M but not in N. For example, consider the flow induced by the vector field

$$(\dot{x}, \dot{y}) = (x^3 - y^2x, x^2y - y^3) \tag{11}$$

shown in Fig. 5. This has an equilibrium at $(0, 0)$ that has the open basin of attraction given by $y^2 > x^2$. However, all points with $y^2 < x^2$ are repelled away to infinity. Thus the origin is an attractor in $\mathbb{R}^2$ but has trivial basin of attraction in the invariant subspace given by the x-axis. In this example this is a degeneracy caused by non-hyperbolicity of the fixed point at the origin. We expect that this behaviour cannot occur in sufficiently hyperbolic systems.

In applications where noise is highly directional, riddled basin attractors may have a degree of sensitivity of an attractor to anisotropic noise that is dependent on whether the noise is in directions in which the basin is unriddled or not as we will see later.

4.2 Internal and Other Riddling Transitions

Given any attractor A with a riddled basin in a system with several invariant subspaces, we can characterize this basin by examining the dimension $\tilde{d}$ of the

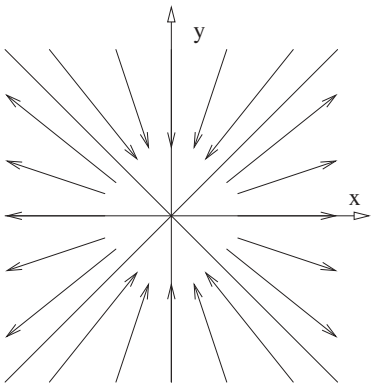


Fig. 5. The dynamics near the non-hyperbolic fixed point at the origin $(0,0)$ for (11) such that it is a Milnor attractor for the full system but not for the system restricted to the invariant subspace $(x,0)$

largest invariant subspace N containing A such that the basin of $\mathcal{B}(A) \cap N$ is unriddled. Note that $\dim(A) \leq \tilde{d} \leq \dim(M)$. Any change in $\tilde{d}$ we term an *internal riddling transition*. It is clear that such transitions will occur in higher-dimensional systems; as A loses asymptotic stability in more directions the index $\tilde{d}$ will decrease.

Similarly one can apply the standard riddling bifurcation criteria in each subspace to predict parameter values when internal riddling transitions will occur; see [16, 48, 51, 63, 81]. Such transitions will typically be rather unclear on varying system parameters if the chaotic attractors are not structurally stable; only by examining normal parameters of the system such that the dynamics on the attractor is left unchanged can one hope to find internal riddling transitions appearing as codimension one transitions. Similar transitions where riddling bifurcations are replaced by blowout bifurcations will also occur in such systems; see for example the systems studied in [5].

4.3 A Numerical Example

If we examine the map (8) for parameter values that are intermediate between strong and weak coupling one may find a wide variety of attractors of different symmetries that are multi-stable. One can also find, for example, (i) chaotic saddles that have basins that are riddled within certain invariant subspaces (ii) attractors that are riddled in some directions but not others and hence (iii) anisotropic bubbling response to anisotropic noise.

Here we consider an attractor in

$$\mathrm{Fix}(S_2 \times S_2) = \{(p,p,q,q) \ : \ p,q \in \mathbb{R}\}.$$

that occurs when $a = 1.71$ and $\epsilon = 0.15$. Figure 6(i) shows a time series on this attractor for an initial condition very close to

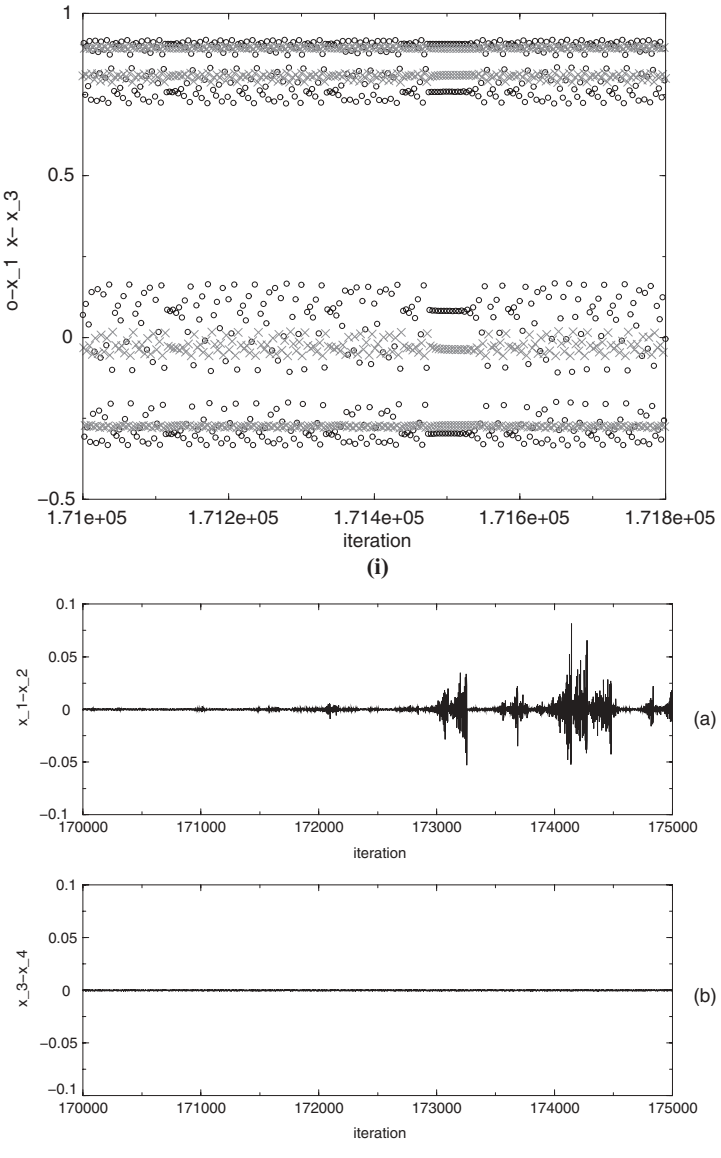


Fig. 6. Anisotropic bubbling behaviour caused by anisotropic riddling of the basin of an attractor for a system of four globally coupled maps (8) that lies in an invariant subspace with symmetry $S_2 \times S_2$. (i) shows time-series for the noise-free system; the circles show the values of x_1 and x_2 while the crosses show the values of x_3 and x_4. In (ii) the attractor in (i) is subject to very low amplitude noise. In (iia) the noise is added in only the x_1 direction giving rise to bubbling in this direction. In (iib) it is added in only the x_3 direction giving a stable response, due to the attractor not being riddled in this direction

$$(p, p, q, q) = (0.00292, 0.00292, 0.8004, 0.8004)$$

and no added noise; $\sigma_j = 0$ for all j. On addition of noise (iia) shows $x_1 - x_2$ for the same initial condition but with $\sigma_1 = 10^{-5}$. The large deviations away from $x_1 = x_2$ indicative of bubbling in the direction $(1, 0, 0, 0)$ and hence indicate that the basin of attraction of this attractor is riddled to perturbations into the invariant subspace (p, q, r, r). By contrast, (iib) shows $x_3 - x_4$ for the same initial condition but with $\sigma_3 = 10^{-5}$. In this case there is apparently stable response indicating that the basin is not riddled into the invariant subspace (p, p, q, r). Note that (i) shows that the statistics of x_1 and x_3 are quite different; the natural invariant measure associated with this attractor is not invariant under the transformation

$$(x_1, x_2, x_3, x_4) \mapsto (x_3, x_4, x_1, x_2).$$

If this symmetry did leave the attractor invariant then by applying Theorem 4.2 either both or neither of the directions $(1, 0, 0, 0)$ and $(0, 0, 1, 0)$ would be riddled. Numerical calculations indicate that the transverse L.E. in the direction $(1, 0, 0, 0)$ is approximately -7.8×10^{-4} whereas it is approximately -0.2842 in the direction $(0, 0, 1, 0)$; this agrees with the anisotropic bubbling observations.

5 Some Open Problems

In the final section we highlight some open problems and themes of interest related to riddled basins.

5.1 Unfolding of Blowout and Riddling

One of the most intriguing aspects of blowout bifurcation and associated transitions to riddled basins are the observation, noted first by [64] in many examples one can clearly classify the nonlinear stability of the system into one of two scenarios: either subcritical, where a riddled basin attractor loses its basin at blowout to become a repellor, or supercritical, where an attractor is locally but not globally riddled before blowout, and after blowout there is on-off intermittency [37, 69, 70] or a stuck-on attractor [3]. (See [67, 68, 84] for more early work on loss of chaotic synchronization)

By analogy with bifurcation of fixed points, it would be nice to understand precisely what determines criticality at blowout and hence whether basins of attraction are locally or globally riddled. The map (1,2,3) was studied in [6] where it was noted that for this map, the two scenarios could be distinguished subcritical (resp. supercritical) if the *essential basin of attraction* of A was zero (resp. positive) measure at the point of blowout.[4] In the general case this

[4] The essential basin of A is defined as the set of points y whose trajectories visit any neighbourhood of A with positive frequency; i.e. such that $\limsup_{n\to\infty} \frac{1}{n}\#\{0 \le k < n \ : \ f^k(y) \in U\} > 0$ for all U neighbourhoods of A.

amounts to a conjecture. Other work has examined *absorbing areas* [21, 52] and weak attractors [16] to give criteria for supercriticality; for example in [16] the presence of a larger weak attractor containing a riddled basin attractor seems to indicate that blowout will be supercritical.

Generally, can one find conditions that ensure that a "scenario" occurs at blowout and if we can do this, find conditions so that the classification of the blowout into a scenario is computable in some sense? So far, it seems that is may only be possible to answer this in a very limited sense; if so, why can one nevertheless observe this scenario in many simulations?

5.2 Riddling for Infinite Dimensions

Many mathematical models that arise include temporal delays and/or spatial extension in such a way that the phase space on which the dynamics occurs is unavoidably infinite dimensional. So far there is no corresponding concept for riddled basin, or indeed Milnor attractor, that works in cases where a nice background measure such as Lebesgue is not available. It should be possible to use ideas such as prevalence [39] in such cases to make some progress but this has not yet been done to my knowledge.

5.3 Pseudo-Riddled and δ-Riddled Basins

A strange feature of riddling is that in cases that one can prove it exists it is often hard to find, while in cases where can prove it does not exist, numerics often seem to say the opposite. For example, in [49] examples of pseudo-riddled basins are presented. In [12] the concept of δ-riddling was used to give a numerical profile of pseudo-riddling by saying that a set is δ-riddled if (6) holds for a given δ; the set is then riddled in the usual sense if it is δ-riddled for all $\delta > 0$. It would be nice to gain a better theoretical understanding of these effects and their consequences, though any simple minded approach clearly has the disadvantage of not being invariant under coordinate changes.

5.4 Genericity of Riddled Basins

In dissipative dynamics that has no imposed invariant subspaces, it is the belief of the writer (and others) that riddled basin attractors do not arise except in exceptional circumstances, namely when there are invariant subspaces forced by system symmetries or for example by other constraints [72]. Can one prove any meaningful results in this direction? An important part of this questions is to understand the prevalence of the appearance of chaotic attractors; see for example [20, 22, 40, 50, 57].

The appearance of a minimal Milnor attractor that has a partially riddled basin was suggested in [16] to be a degenerate case; can one prove a sense in which this is for example non-generic in a class of smooth systems?

5.5 Unstable Attractors

There has been some very interesting recent work looking at dynamical systems of globally delay pulse coupled oscillator systems, motivated by simple models of neural systems. These systems can possess attractors that are an extreme case of riddling; there is a neighbourhood of the attractor that has zero measure intersection with the attractor's basin. Such *unstable attractors* have been found to be quite widespread in the dynamics of certain systems [80]. It is a challenge to understand constraints on the appearance of such attractors and their characteristic properties.

5.6 Non-Ergodic Attractors and Riddling

In most standard examples where one can prove the existence of riddled basins the attractor in question is ergodic, namely it possesses an SRB measure whose support is the attractor. Nonetheless, in the presence of symmetries it has been recognised for some time that for many cases one can find structurally stable attractors that are not ergodic; in particular robust attracting heteroclinic cycles [35, 46, 47] between saddle equilibria, periodic orbits or even chaotic saddles [4, 27, 31]. These attractors are non-ergodic and not even transitive (no dense orbits); they consist of chains of connecting orbits and invariant sets [13, 14].

As a particular example of this in a finite coupled map lattice is the system with one-directional coupling considered in [15]:

$$X_{n+1}^k = f(X_n^k)\, e^{-\gamma X_n^{k-1}}$$

where $n \in \mathbb{N}$, $k \in \mathbb{Z}$ is taken modulo 3 and $f(x) = rx(1-x)$. Due to the presence of invariant subspaces $X^k = 0$ (for any k) if we choose $2 < r < 4$ and γ large enough this system can be shown to have open sets of spatially periodic initial conditions that are attracted to states with no convergence of ergodic averages; see [15] for more details.

For a skew product system of a chaotic attractor forcing a robust heteroclinic cycle one can exhibit examples where the cycle has a locally riddled basin even though the attractor is non-ergodic [4]. How general is this?

5.7 Other Problems Related to Riddled Basins

Some researchers have implied that there are connections between the computability of riddled basins and decidability [65, 76]. It would be of interest to know if riddling may help one to understand the nature of chaotic behaviour in area-preserving maps such as the standard map, or in other area-preserving maps; for example Fig. 7 shows in black the invariant set of points that approach arbitrarily close the discontinuity for an area-preserving map that arises in signal processing; see [10, 12] for a discussion of this and generalization of the results such as Theorem 2.1 to area-preserving discontinuous

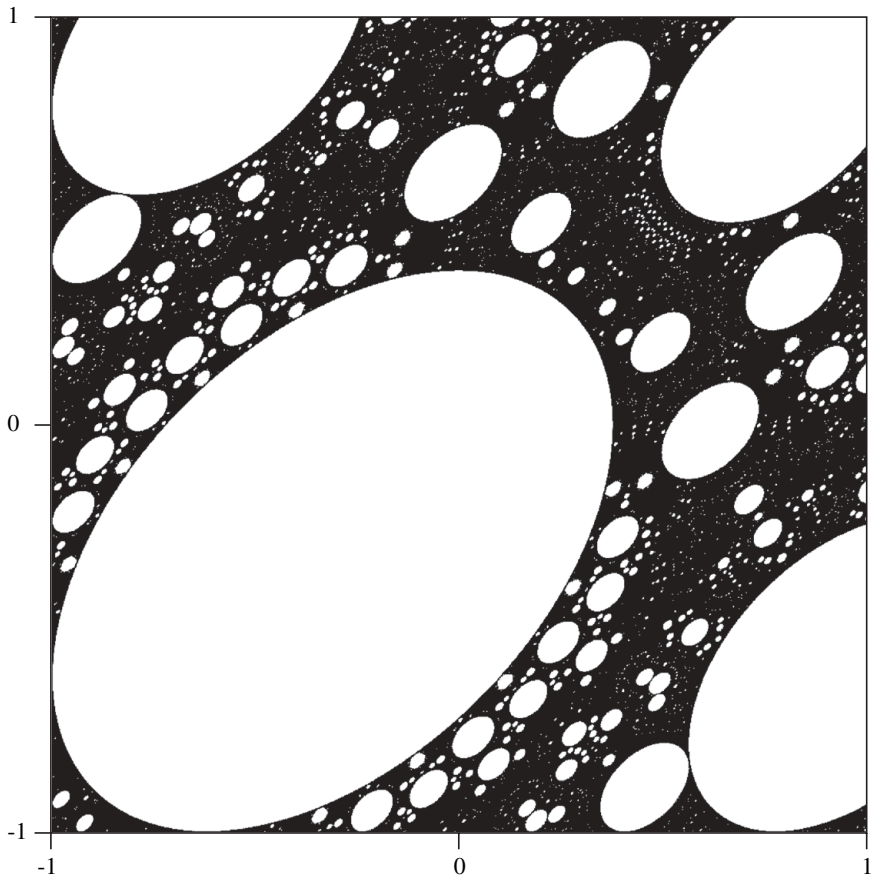


Fig. 7. The *black* region shows the set of points within the square $[-1, -7/8]^2$ whose orbits accumulate on the discontinuity for the area-preserving map $(x, y) \mapsto (y, g(-x + 0.9y))$ where $g(x) = x$ for $x \in [-1, 1)$ and $g(x + 2) = g(x)$. It is an open question whether this set is riddled; the white regions are packed with invariant curves centred on periodic points and the map can be viewed as a planar piecewise isometry

maps. It would be of great interest to be able to determine if the set shown in this figure is riddled or not. Other problems include getting a better understanding of properties of intermingled basins.

Acknowledgements

I would like to thank many collaborators over the years who have helped and stimulated my understanding of riddling effects. In particular I would like to thank P. Aston, M. Breakspear, J. Buescu, E. Covas, M. Field,

X. Fu, I. Melbourne, M. Nicol, R. Roy, I. Stewart, R. Tavakol and J. Terry. I would also like to thank J.-R. Chazottes and B. Fernandez for organizing the school-forum CML2004 that provided the occasion to prepare these notes.

References

1. J.C. Alexander, J.A. Yorke, Z. You and I. Kan,: Riddled Basins. *Int. Journal of Bifurcations and Chaos* **2** (1992), 795–813. 193, 197
2. J.C. Alexander, B.R. Hunt, I. Kan, and J.A. Yorke, Intermingled basins for the triangle map Ergodic Theory Dynam. Systems **16** (1996), no. 4, 651–662. 193
3. P. Ashwin: Attractors stuck on to invariant subspaces *Phys. Lett. A* **209**, 338–344 (1995). 200
4. P. Ashwin: Cycles homoclinic to chaotic sets; robustness and resonance *Chaos* **7**, 207–220 (1997). 202
5. P. Ashwin and P.J. Aston: Blowout bifurcations of codimension two. *Physics Letters A* **244**, 261–270 (1998). 198
6. P. Ashwin, P.J. Aston, P. and M. Nicol: On the unfolding of a blowout bifurcation. *Physica D* **111**, 81–95 (1998). 183, 184, 185, 200
7. P. Ashwin and M. Breakspear: Anisotropic properties of riddled basins. *Phys. Lett. A* **280**, 139–145 (2001). 181, 193, 195
8. P. Ashwin, J. Buescu and I.N. Stewart: Bubbling of attractors and synchronisation of oscillators. *Phys. Lett. A* **193**, 126–139 (1994). 189, 193
9. P. Ashwin, J. Buescu and I. Stewart: From attractor to chaotic saddle: a tale of transverse instability. *Nonlinearity* **9**, 703–737 (1996). 189, 192, 193, 195, 197
10. P. Ashwin, W. Chambers and G. Petkov: Lossless digital filter overflow oscillations; approximation of invariant fractals. *Int. J. Bifur. Chaos* **7**, 2603–2610 (1997). 202
11. P. Ashwin, E. Covas and R.K. Tavakol: Transverse instability for non-normal parameters. *Nonlinearity* **12**, 563–577 (1999). 193, 195
12. P. Ashwin, X.-C. Fu and J.R. Terry: Riddling and invariance for discontinuous maps preserving Lebesgue measure. *Nonlinearity* **15**, 633-645 (2002). 188, 201, 202
13. P. Ashwin and A.M. Rucklidge: Cycling chaos: its creation, persistence and loss of stability in a model of nonlinear magnetoconvection. *Physica D* **122**, 134–154 (1998). 191, 202
14. P. Ashwin, A.M. Rucklidge and R. Sturman: Cycling attractors of coupled cell systems and dynamics with symmetry. In: *Proceedings of Workshop on Synchronization, Crimea, 2002.* Eds A. Pikovsky and Y. Maistrenko, Kluwer (2003). 191, 202
15. P. Ashwin and R. Sturman: Decelerating defects and non-ergodic critical behaviour in a unidirectionally coupled map lattice. *Phys Letters A* **309**, 423–434 (2003). 202
16. P. Ashwin and J.R. Terry: On riddling and weak attractors. *Physica D* **142**, 87–100 (2000). 183, 187, 188, 189, 193, 198, 201
17. P.J. Aston and M. Dellnitz: Symmetry breaking bifurcation of chaotic attractors. *Int. J. Bif. Chaos* **5**, 1643-1676 (1995). 191
18. V. Baladi and M. Viana: Strong stochastic stability and rate of mixing for unimodal maps *Ann. Sci. École Norm. Sup. (4)* **29**, 483–517 (1996). 190

19. E. Barany, M. Dellnitz and M. Golubitsky: Detecting the symmetry of attractors. *Physica D* **67**, 66–87 (1993). 191, 196
20. E. Barreto, B.R. Hunt, C. Grebogi and J. Yorke: From high dimensional chaos to stable periodic orbits: the structure of parameter space. *Phys. Rev. Lett.* **78**, 4561–4564 (1997). 195, 201
21. G.-I. Bischi and L. Gardini: Role of invariant and minimal absorbing areas in chaos synchronization. *Phys. Rev. E* **58**, 5710–5719 (1998). 201
22. M. Benedicks and L. Carleson: On iterations of $1 - ax^2$ on $[-1, 1]$. *Ann. Math.* (2) **122**, 1–25 (1985). 201
23. J. Buescu: *Exotic Attractors.* (Birkhäuser Verlag, Basel 1997). 192
24. Y. Cao: A note about Milnor attractor and riddled basin. *Chaos, Solitons and Fractals* **19**, 759–764 (2004). 183, 193
25. P. Chossat and M. Golubitsky: Symmetry increasing bifurcation of chaotic attractors, *Physica D* **32**, 423–436 (1988). 191
26. E. Covas, P. Ashwin and R.K. Tavakol: Non-normal parameter blowout bifurcation in a truncated mean field dynamo. *Phys. Rev. E* **56**, 6451–6458 (1997). 195
27. M. Dellnitz, M. Field, M. Golubitsky, A. Hohmann and J. Ma: Cycling chaos, *IEEE Trans. Circuits and Systems-I.* **42**, 821–823 (1995). 191, 202
28. M. Ding and W. Yang: Observation of intermingled basins in coupled oscillators exhibiting synchronized chaos. *Phys. Rev. E* **54**, 2489–2494 (1996). 193, 194
29. J.-P. Eckmann and D. Ruelle: Ergodic Theory of Chaos and Strange Attractors. *Rev. Mod. Phys.* **57**, 617–656 (1985). 182, 192
30. J.D. Farmer: Sensitive dependence on parameters in nonlinear dynamics. *Phys. Rev. Lett.* **55**, 351–354 (1985). 187
31. M. Field. *Lectures on Dynamics, Bifurcation and symmetry.* Pitman Research Notes in Mathematics, **356**, Pitman (1996). 191, 202
32. K. Falconer. *Fractal Geometry* (John Wiley & sons, Chichester 1990). 187, 188
33. M. Golubitsky and I. Stewart. *The symmetry perspective* (Birkhäuser Verlag, Boston 2002). 191
34. C. Grebogi, S.W McDonald, E. Ott & J.A. Yorke. Exterior dimensions of fat fractals *Phys. Lett. A* **110**:1–4 (1985). 187, 188
35. J. Guckenheimer and P. Holmes. Structurally stable heteroclinic cycles, *Math. Proc. Camb. Phil. Soc.* **103**:189–192 (1988). 202
36. M. Hasler, Y. Maistrenko and O. Popovych: Simple example of partial synchronization of chaotic systems. *Phys. Rev. E* **58**, 6843–6346 (1998). 193
37. J.F. Heagy, N. Platt and S.M. Hammel: Characterization of on-off intermittency. *Phys. Rev. E* **49**, 1140–1150 (1994). 200
38. J.F. Heagy, T. Carroll and L. Pecora: Experimental and numerical evidence for riddled basins in coupled chaotic systems. *Phys. Rev. Lett.* **73**, 3528–31 (1994). 193
39. B.R. Hunt: The Prevalence of Continuous Nowhere Differentiable Functions *Proc. Amer. Math. Soc.* **122**, 711–717 (1994). 201
40. M. Jakobson: Absolutely continuous invariant measures for one parameter families of one-dimensional maps, *Comm. Math. Phys.* **81**, 39–88 (1981). 201
41. I. Kan: Open sets of diffeomorphisms having two attractors, each with an everywhere dense basin. *Bull. Am. Math. Soc.* **31**, 68–74 (1994). 193
42. K. Kaneko: On the strength of attractors in a high-dimensional system: Milnor attractor network, robust global attraction, and noise-induced selection. *Physica D* **124**, 322–344 (1998). 191
43. K. Kaneko: Mean field fluctuation of a network of chaotic elements. *Physica D* **55**, 368–384 (1992).

44. J.L. Kaplan and J.A. Yorke: Chaotic behaviour of multidimensional difference equations. In *Functional differential equations and approximations of fixed points*, (H.-O. Peitgen and H.-O. Walther, eds) Springer Lecture Notes in Maths **730** (1979). 189
45. L. Kocarev, A. Shang and L.O. Chua: Transitions in dynamical regimes by driving: a unified method of control and synchronisation of chaos. *Intl. J. Bifurcation and Chaos* **3**, 479–483 (1993).
46. M. Krupa and I. Melbourne: Asymptotic stability of heteroclinic cycles in systems with symmetry, *Erg. Th. Dyn. Sys.* **15**, 121–147 (1995). 202
47. M. Krupa: Robust heteroclinic cycles, *Journal of Nonlinear Science* **7**, 129–176 (1997). 202
48. Y.-C. Lai, C. Grebogi, J.A. Yorke and S.C. Venkataramani: Riddling bifurcation in chaotic dynamical systems. *Phys. Rev. Lett.* **77**, 55–58 (1996). 198
49. Y.-C. Lai: Pseudo-riddling in chaotic systems. *Phys. D* **150**, 1–13 (2001). 201
50. A. Lasota and J.A. Yorke: On the existence of invariant measures for piecewise monotonic transformations. *Trans. Am. Math. Soc.* **186**, 481–488 (1973). 201
51. Yu.L. Maistrenko, V.L. Maistrenko, A. Popovich and E. Mosekilde: Transverse instability and riddled basins in a system of two coupled logistic maps. *Phys. Rev. E* **57**, 2713–2724 (1998). 198
52. Yu.L. Maistrenko, V.L. Maistrenko, A. Popovich and E. Mosekilde: Role of the absorbing area in chaotic synchronization. *Phys. Rev. Lett.* **80**, 1638–1641 (1998). 201
53. R. Mañé: *Ergodic Theory and Differentiable Dynamics.* (Springer-Verlag, Berlin 1987). 192
54. I. Melbourne: An example of a non-asymptotically stable attractor. *Nonlinearity* **4**, 835–844 (1991). 187
55. I. Melbourne, M. Dellnitz and M. Golubitsky: The Structure of Symmetric Attractors, *Arch. Rational. Mech. Anal.* **123**, 75–98 (1993). 191
56. J. Milnor: On the concept of attractor. *Commun. Math. Phys.* **99**, 177–195 (1985); Comments *Commun. Math. Phys.* **102**, 517–519 (1985). 182, 183
57. L. Mora and M. Viana: Abundance of strange attractors. *Acta Math.* **171**, 1–71 (1993). 201
58. E. Mosekilde, Y. Maistrenko and D. Postnov: *Chaotic synchronization. Applications to living systems.* (World Scientific Publishing, River Edge NJ, 2002). 181
59. H. Nakajima and Y. Ueda: Riddled basins of the optimal states in learning dynamical systems. *Physica D* **99**, 35–44 (1996). 193
60. V.I. Oseledec: A multiplicative ergodic theorem: Liapunov characteristic numbers for dynamical systems. *Trans. Mosc. Math. Soc.* **19**, 197–231, (1968). 192
61. E. Ott: *Chaos in Dynamical Systems* (Second edition) (Cambridge University Press 2002). 182
62. E. Ott, J.C. Sommerer, J. Alexander, I. Kan and J.A. Yorke: Scaling behaviour of chaotic systems with riddled basins. *Phys. Rev. Lett.* **71**, 4134–4137 (1993). 188, 189
63. E. Ott, J.C. Sommerer, J. Alexander, I. Kan and J.A. Yorke: A transition to chaotic attractors with riddled basins. *Physica D* **76**, 384–410 (1994). 198
64. E. Ott and J.C. Sommerer: Blowout bifurcations: the occurrence of riddled basins and on-off intermittency. *Phys. Lett. A* **188**, 39–47 (1994). 200
65. M.W. Parker: Undecidability in $\mathbb{R}^n$; riddled basins, the KAM tori and the stability of the solar system. *Philos. Sci.* **70**, 359–382 (2003). 202
66. Y.B. Pesin: Characteristic Liapunov exponents and smooth ergodic theory. *Russian Math. Surv.* **32**, 55–114 (1978). 192

67. A.S. Pikovsky: On the interaction of strange attractors. *Z. Phys. B*, **55**, 149–154 (1984). 200
68. A.S. Pikovsky and P. Grassberger: Symmetry breaking of coupled chaotic attractors. *J. Phys. A*, **24**, 4587–4597 (1991). 200
69. N. Platt, E. Spiegel and C. Tresser: On-off intermittency; a mechanism for bursting. *Phys. Rev. Lett.* **70**, 279–282 (1993). 200
70. N. Platt, S.M. Hammel and J.F. Heagy: Effects of additive noise on on-off intermittency. *Phys. Rev. Lett.* **72**, 3498–3501 (1994). 200
71. C. Pugh and M. Shub: Ergodic attractors. *Trans. Am. Math. Soc.* **312**, 1–54 (1989). 182, 192
72. D.A. Rand, H. Wilson and J. McGlade: Dynamics and evolution: evolutionarily stable attractors, invasion exponents and phenotype dynamics. *Phil. Trans. R. Soc. Lond. B*, **343**, 261–283 (1994). 201
73. D. Ruelle: Analyticity properties of the characteristic exponents of random matrix products, *Advances in Math.* **32**, 68–80 (1979). 195
74. D. Ruelle: *Chaotic evolution and strange attractors*, (Lezione Lincee, Cambridge University Press, 1989). 195
75. R.J. Sacker and G.R. Sell: The spectrum of an invariant submanifold. *J. Diff. Equations* **38**, 135–160 (1980). 192
76. A. Saito and K. Kaneko: Inaccessibility and indecidability in computation, geometry and dynamical systems. *Physica D* **155**, 1–33 (2001). 202
77. S.H. Strogatz: *Synch: the emerging science of spontaneous order*, (Hyperion Press 2003). 181
78. K. Sygmund: Generic properties of invariant measures for Axiom A diffeomorphisms. *Inv. Math.* **11**, 99–109 (1970). 193
79. A.V. Taborev, Y.L. Maistrenko, and E. Mosekilde: Partial synchronization in a system of coupled logistic maps. *Int. J. Bifurcation and Chaos* **10**, 1051–1066 (2000). 191
80. M. Timme, F. Wolf, and T. Geisel: Prevalence of unstable attractors in networks of pulse-coupled oscillators, *Phys. Rev. Lett.* **89**, 154105 (2002). 202
81. S.C. Venkataramani, B.R. Hunt, E. Ott, D.J. Gauthier and J.C. Bienfang: Transitions to bubbling of chaotic systems. *Phys. Rev. Lett.* **77**, 5361–5364 (1996). 198
82. R. L. Wheeden and A. Zygmund: Measure and integral. (Marcel Dekker Inc, New York 1977). 188
83. M. Woltering and M. Markus: Riddled basins of coupled elastic arches. *Phys. Lett. A* **260**, 453–461 (1999). 193
84. H. Yamada and T. Fujisaka: Stability theory of synchronised motion in coupled-oscillator systems. *Prog. Theor. Phys.* **70**, 1240–1248 (1984). 200

A Spectral Gap for a One-dimensional Lattice of Coupled Piecewise Expanding Interval Maps

G. Keller[1] and C. Liverani[2]

[1] Mathematisches Institut, Universität Erlangen-Nürnberg, Bismarckstr. 1 1/2, 91054 Erlangen, Germany
keller@mi.uni-erlangen.de

[2] Dipartimento di Matematica, Università di Roma II (Tor Vergata), Via della Ricerca Scientifica, 00133 Roma, Italy
liverani@mat.uniroma2.it

Abstract. We study one-dimensional lattices of weakly coupled piecewise expanding interval maps as dynamical systems. Since neither the local maps need to have full branches nor the coupling map needs to be a homeomorphism of the infinite dimensional state space, we cannot use symbolic dynamics or other techniques from statistical mechanics. Instead we prove that the transfer operator of the infinite dimensional system has a spectral gap on suitable Banach spaces generated by measures with marginals that have densities of bounded variation. This implies in particular exponential decay of correlations in time and space.

1 Introduction

Typical dynamical systems have a multitude of invariant probability measures. There are essentially two ways to characterize the "physically relevant" ones among them: in the spirit of statistical mechanics one can look at those measures which satisfy a variational principle with a potential of the type "logarithm of the unstable Jacobian". From a more dynamical perspective one may look at those measures which are absolutely continuous w.r.t. the natural volume measure m on the state space, or at those for which the space averages of regular observables equal the corresponding time averages for a set of initial conditions of positive m-measure. In many cases both approaches lead to the same result. In the case of coupled map lattices, which are infinite-dimensional dynamical systems, one needs some extra care to apply these ideas. For the statistical mechanics approach this is done in other chapters of this book. Here we concentrate on the dynamical systems approach.

Let L be a finite or countable index set, e.g. $L = \mathbb{Z}$ or $L = \mathbb{Z}/d\mathbb{Z}$ and let $I = [0, 1]$. We investigate time-discrete dynamics on the state space $X = I^L$, composed of independent chaotic actions on each component I of X and of some weak interaction between the components that does not destroy the

G. Keller and C. Liverani: *A Spectral Gap for a One-dimensional Lattice of Coupled Piecewise Expanding Interval Maps*, Lect. Notes Phys. **671**, 115–151 (2005)
www.springerlink.com

chaotic character of the whole system. More specifically, let τ be a piecewise C^2 map from I to I with singularities at $\zeta_1, \dots, \zeta_{N-1} \in (0,1)$ in the sense that τ is monotone and C^2 on each component of $I \setminus \{\zeta_0 = 0, \zeta_1, \dots, \zeta_{N-1}, \zeta_N = 1\}$. We assume that $\tau''/(\tau')^2$ is bounded and that τ satisfies the following combined expansion and regularity assumption:

There is $M \in \mathbb{N}$ such that $\kappa_M := \inf |(\tau^M)'| > 2$ and such that

$$\tau^m(0), \tau^m(\zeta_1 \pm), \dots, \tau^m(\zeta_{N-1} \pm), \tau^m(1) \notin \{\zeta_1, \dots, \zeta_{N-1}\} \tag{1}$$

for $m = 0, \dots, M-1$.

If $\inf |\tau'| > 2$ this condition is trivially satisfied for $M = 1$. A simple but prominent example of such a map with $M > 1$ is a symmetric mixing tent map, i.e. a map $\tau_s(x) = s(\frac{1}{2} - |x - \frac{1}{2}|)$ with slope $s \in (\sqrt{2}, 2]$. It satisfies (1) with $N = 2$, $\zeta_1 = \frac{1}{2}$, $M = 2$ and $\kappa_M = s^2$.[1]

Now a map $T_0 : X \to X$ describing the uncoupled dynamics is defined by

$$(T_0 \boldsymbol{x})_i = \tau(x_i) \quad (i \in L) \tag{2}$$

and coupled maps $T_\epsilon := \Phi_\epsilon \circ T_0$ are introduced using appropriate continuous couplings $\Phi_\epsilon : X \to X$ close to the identity on X.[2] One of the most widely used couplings in numerical studies – which despite its simplicity resisted for quite some time a rigorous mathematical treatment – is the diffusive nearest neighbor coupling on $\mathbb{Z}$ or $\mathbb{Z}/d\mathbb{Z}$

$$(\Phi_\epsilon \boldsymbol{x})_i = \frac{\epsilon}{2} x_{i-1} + (1-\epsilon) x_i + \frac{\epsilon}{2} x_{i+1} \quad (i \in L)\,. \tag{3}$$

It is an example of a class of more general C^2-couplings Φ_ϵ whose C^2 distance to the identity Φ_0 is of order ϵ and is controlled in terms of constants $a_1, a_2 > 0$ – see Sect. 3.1 for details. We say that such a coupling has *finite coupling range* if there is $w > 0$ such that $\partial_j \Phi_{\epsilon,i} = 0$ whenever $|i - j| > w$.

Our main result is:

Theorem 1.1. *Let $L = \mathbb{Z}$. Given a mixing[3] local map τ as introduced above and given $a_1, a_2, w > 0$, there exists $\epsilon_{\max} > 0$ such that for each (a_1, a_2)-coupling Φ_ϵ with coupling range w and each $\epsilon \in [0, \epsilon_{\max}]$ holds:*

[1] An elementary discussion of the basic dynamical properties of these maps can be found in [1]. For the mixing property see [2].

[2] The regularity assumption in (1) seems unavoidable if a weakly coupled system T_ϵ is to behave like a small perturbation of T_0, because weak couplings affect each individual map τ like a small perturbation, and it is known that in the absence of the above assumption arbitrarily small perturbations can change the dynamics of τ completely, see the examples in [3, 4, 5].

[3] Under the assumptions made on τ there exists at least one invariant probability density for τ. We say that τ is *mixing*, if no power of τ has any other invariant probability density. This will be discussed in some detail in Sect. 2.5.

1. *The coupled system T_ϵ has an invariant probability measure μ_ϵ whose finite-dimensional marginals are absolutely continuous w.r.t. Lebesgue measure and have densities of bounded variation. It has finite entropy density, and it is unique among all measures from this class for which the variation of the marginals increases at most subexponentially with the dimension.*
2. *There are constants $\gamma, \gamma', \theta \in (0,1)$ and $C, C' > 0$ such that for bounded observables $\phi, \psi : X \to \mathbb{R}$ which depend only on coordinates $x_{a+1}, \dots, x_b$,*

$$\left| \int \phi \cdot (\psi \circ T_\epsilon^n)\, d\mu_\epsilon - \int \phi\, d\mu_\epsilon \int \psi\, d\mu_\epsilon \right| \le C\theta^{-(b-a)}\, \gamma^n\, \|\phi\|_{C^1} \|\psi\|_{C^0} \quad (4)$$

and

$$\left| \int \phi \cdot (\psi \circ \sigma^n)\, d\mu_\epsilon - \int \phi\, d\mu_\epsilon \int \psi\, d\mu_\epsilon \right| \le C'\gamma'^{|n|-(b-a)}\, \|\phi\|_{C^0} \|\psi\|_{C^0} \quad (5)$$

where σ is the left shift on $X = I^{\mathbb{Z}}$.
3. *The distance (in a suitable metric) between μ_ϵ and μ_0 is of order $\epsilon \ln \epsilon^{-1}$.*

The proof of this theorem relies on a spectral analysis of the transfer operator associated with T_ϵ. It combines results and ideas from [6, 7, 8, 9] and is developed step by step in this chapter. The existence part and the finiteness of the entropy density are proved in Theorem 4.1 in Sect. 4.4. Uniqueness of μ_ϵ, the exponential decay of correlations, and the estimate on the distance between μ_ϵ and μ_0 are derived in Sects. 4.7, 4.8, and 4.9 from Theorem 4.3, which guarantees the existence of a spectral gap for the transfer operator of T_ϵ on suitable Banach spaces. In Sect. 4.8 we also prove the following strong law of large numbers (compare [10, Theorem 5.1]):

Corollary 1.1. *In the situation of Theorem 1.1, let $\psi : X \to \mathbb{R}$ be a continuous observable. Let $f : I \to \mathbb{R}$ be any probability density of bounded variation, and let (fm) be the corresponding probability measure on I. Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \psi(T_\epsilon^k(\boldsymbol{x})) = \int \psi\, d\mu_\epsilon \quad \text{for } (fm)^{\mathbb{Z}}\text{-a.e. } \boldsymbol{x} \quad (6)$$

where $(fm)^{\mathbb{Z}}$ is the infinite product measure on $X = I^{\mathbb{Z}}$ with one-dimensional factors (fm).

This result suggests the interpretation of μ_ϵ as the unique physical (or observable) measure of T_ϵ.[4] It is supported by the stability of μ_ϵ under independent random noise discussed (without proof) in Sect. 4.9.

[4] For a discussion of physical measures and related notions in more general settings see e.g. the contribution of L.A. Bunimovich in this volume.

2 Dynamics at a Single Site

The dynamics at each single site of our system is modeled by a mixing piecewise expanding map. We will see that the dynamics of such maps are statistically stable in several respects: there is a unique stationary probability density towards which each initial density converges under the action of the dynamics (*asymptotic stability*), and neither this stationary density nor the rate at which initial densities are attracted by it change much under small perturbations of the map. Therefore we can expect that in weakly coupled systems, where the mutual coupling between the single site maps can also be interpreted as a kind of perturbation, the behavior of the system at single sites does not change drastically under the influence of the coupling.

2.1 Piecewise Expanding Maps

We say that a map $\tau : I \to I$ is *piecewise expanding (p.w.e.)* if

- there are $\zeta_1, \dots, \zeta_{N-1} \in (0,1)$ which define subintervals $I_i = (\zeta_{i-1}, \zeta_i)$ such that each $\tau|_{I_i}$ is monotone and uniformly C^2, and
- there are $M \in \mathbb{N}$ and $\kappa_M > 2$ such that $|(\tau^M)'| \geq \kappa_M$.[5]

Our assumptions imply that $D_m := \sup |(\frac{1}{(\tau^m)'})'| < \infty$ for all $m \in \mathbb{N}$.

2.2 The Transfer Operator

As the derivatives of p.w.e. maps grow exponentially, the trajectory-wise dynamics are very sensitive to initial conditions. But at the same time this instability is responsible for good asymptotic properties of the *transfer operator*[6] P_τ which describes the evolution of initial densities under the dynamics. This operator associates to each measurable $f : I \to \mathbb{R}$ the function $P_\tau f : I \to \mathbb{R}$,

$$P_\tau f(x) = \sum_{i=1}^{N} \left(\frac{f}{|\tau'|} \right) \circ (\tau|_{I_i})^{-1}(x) \cdot 1_{\tau(I_i)}(x) \tag{7}$$

where $1_{\tau(I_i)}$ denotes the indicator function of the set $\tau(I_i)$. By change of variables it follows that for Lebesgue (m) integrable $f : I \to \mathbb{R}$ and bounded measurable $\psi : I \to \mathbb{R}$

$$\int_I P_\tau f(x)\, \psi(x)\, dx = \int_I f(x)\, \psi(\tau(x))\, dx\ . \tag{8}$$

[5] This means that $|(\tau^M)'(x)| \geq \kappa_M$ at all points x where this derivative is defined, i.e. at all x such that $\tau^i(x) \notin \{\zeta_0, \dots, \zeta_N\}$ for $i = 0, \dots, M-1$. In the sequel all expressions involving derivatives of τ should be read in this way. Note that it suffices to have $|(\tau^m)'| > \kappa_m > 1$ for some $m \in \mathbb{N}$ and to choose $M = km$ such that $\kappa_m^k > 2$.

[6] also called *Perron–Frobenius operator*

In particular, if $f = \tilde{f}$ m-a.e., then also $P_\tau f = P_\tau \tilde{f}$ m-a.e. so that P_τ can be interpreted as an operator on L^1_I, the space of equivalence classes of Lebesgue integrable functions from I to $\mathbb{R}$. As an operator on L^1_I, P_τ is unambiguously defined by (8). The following properties of P_τ are elementary consequences of its definition (see e.g. [11, Sect. 4.2]):

$$P_\tau \text{ is linear and positive.} \tag{9}$$

$$\textstyle\int_I P_\tau f \, dm = \int_I f \, dm \text{ and } \int_I |P_\tau f| \, dm \le \int_I |f| \, dm \text{ for all } f \in L^1_I. \tag{10}$$

$$\begin{aligned} &P_\tau h = h \text{ if and only if } \mu = hm \text{ is a } \tau\text{-invariant measure, i.e. if} \\ &\textstyle\int f \circ \tau \, d\mu = \int f \, d\mu \text{ for each bounded measurable } f : I \to \mathbb{R}. \end{aligned} \tag{11}$$

$$P_{\tau_2 \circ \tau_1} = P_{\tau_2} P_{\tau_1} \text{ whenever the three operators are well defined.} \tag{12}$$

Remark 2.1. The "(pre-)dual" characterization of transfer operators by (8) and its elementary consequences (9)–(12) are not special for 1D maps. They are valid with exactly the same proofs in rather abstract settings, see e.g. [12, Sect. 3.2]. Therefore we will use them in later sections where we study transfer operators for systems of maps without recalling them in detail.

2.3 Functions of Bounded Variation

In order to guarantee the existence of a unique invariant density and its asymptotic stability one needs to study how P_τ acts on spaces of more regular functions. The first space that comes to mind is probably $C^1(I)$, but as $P_\tau f$ may have discontinuities even if f has none (see the explicit formula (7) for $P_\tau f$), this space is not invariant under P_τ. The next natural choice that preserves as much of the flavor of $C^1(I)$ but allows for discontinuities is the space $BV(I)$ of functions of bounded variation.

The variation of a C^1-function $f : I \to \mathbb{R}$ can be defined as

$$V(f) = \int_0^1 |f'(x)| \, dx \, . \tag{13}$$

Approximating this integral by Riemann sums yields the more common expression

$$V(f) = \sup \left\{ \sum_{i=1}^{r} |f(\xi_i) - f(\xi_{i-1})| \right\} \tag{14}$$

where the supremum extends over all finite partitions $0 \le \xi_0 < \xi_1 < \ldots < \xi_r \le 1$ of $[0, 1]$. This expression is well-defined for any measurable $f : I \to \mathbb{R}$. A third characterization follows from the first one in view of the integration by parts formula: let

$$\mathcal{T}_{I,0} = \{\varphi \in C^1(I) : |\varphi| \le 1, \varphi(0) = \varphi(1) = 0\} \tag{15}$$

be a set of C^1-test functions on I bounded by 1. Then

$$V(f) = \sup_{\varphi\in\mathcal{T}_{I,0}} \int_0^1 f'(x)\varphi(x)\,dx = \sup_{\varphi\in\mathcal{T}_{I,0}} \int_0^1 f(x)\varphi'(x)\,dx\,. \tag{16}$$

Just as the previous one this characterization can be used to define the variation of any function $f \in L^1_I$ (and not merely that of C^1-functions). Indeed, (14) leads to the definition

$$\mathrm{var}_I(f) = \inf\{V(\tilde f) : \tilde f = f \ m\text{-a.e.}\} \tag{17}$$

and (16) extends immediately to

$$\mathrm{var}_I(f) = \sup_{\varphi\in\mathcal{T}_{I,0}} \int_I f(x)\varphi'(x)\,dx\,. \tag{18}$$

It is a little extra piece of work to show that the definitions given in (17) and in (18) really coincide.[7] In the sequel we will only use the definition via test functions in (18). Note that

$$\sum_{i=1}^N \mathrm{var}_{\bar I_i}(f) \le \mathrm{var}_I(f)\,. \tag{19}$$

This follows because if $\varphi_i \in \mathcal{T}_{\bar I_i,0}$ $(i = 1,\dots,N)$, then $\varphi : I \to \mathbb{R}$, which is (unambiguously!) defined by $\varphi(x) = \varphi_i(x)$ if $x \in \bar I_i$, belongs to $\mathcal{T}_{I,0}$. A direct consequence of the definition of variation in (18) is

$$\mathrm{var}_I(f) \le \liminf_{n\to\infty} \mathrm{var}_I(f_n) \tag{20}$$

whenever $f, f_n \in L^1_I$ and $\lim_{n\to\infty} \int_I |f - f_n|\,dm = 0$. Here (and in the sequel) we use $\int |f|\,dm$ as a shorthand notation for $\int |f(x)|\,dx$. We denote

$$BV(I) = \{f \in L^1_I : \mathrm{var}_I(f) < \infty\}\,. \tag{21}$$

All these considerations apply to any compact interval I, not just to $I = [0,1]$.

For technical reasons we will often prefer to work with the following variant of the notion of variation. For any compact interval J, let

$$\mathcal{T}_J = \{\varphi \in C^1(J) : |\varphi| \le 1\} \tag{22}$$

and define

$$\mathrm{Var}_J(f) = \sup_{\varphi\in\mathcal{T}_J} \int_J f(x)\varphi'(x)\,dx\,. \tag{23}$$

Here is a first observation on "Var".

[7] See e.g. [11, Theorem 2.3.12].

Lemma 2.1. *Let $J = [a, b]$, and suppose that $(a + \delta, b - \delta)$ is a neighborhood of I for some $\delta > 0$. If $f \in C^1(J)$ with $f(x) = 0$ for $x \notin (a + \delta, b - \delta)$, then*

$$\mathrm{Var}_I(f) \le \int_J |f'|\, dm\ .$$

Proof. Each $\varphi \in \mathcal{T}_I$ can be extended to a function $\tilde{\varphi} : J \to \mathbb{R}$ by linear interpolation between the points $(a|0)$ and $(a+\delta|\varphi(0))$ on the interval $[a, a+\delta]$ and between the points $(b - \delta|\varphi(1))$ and $(b|0)$ on the interval $[b - \delta, b]$, and by the constant values $\varphi(0)$ and $\varphi(1)$ on the intervals $[a + \delta, 0]$ and $[1, b - \delta]$ respectively. In this way, $\tilde{\varphi}$ is continuous, $\sup|\tilde{\varphi}| \le 1$, $\tilde{\varphi}(a) = \tilde{\varphi}(b) = 0$, and $\tilde{\varphi}$ is differentiable except at possibly four points. Hence

$$\int_I f\varphi'\, dm = \int_{a+\delta}^{b-\delta} f\tilde{\varphi}'\, dm = \int_a^b f\tilde{\varphi}'\, dm = -\int_a^b f'\tilde{\varphi}\, dm \le \int_a^b |f'|\, dm$$

from which the lemma follows. □

The next lemma is a kind of tool-box for our work with "var" and "Var".

Lemma 2.2. *Let $J = [a, b]$, $f \in L^1_J$, $\dot{\varphi} \in L^\infty_J$, $c \in \mathbb{R}$, and let $\varphi : J \to \mathbb{R}$, $\varphi(x) = c + \int_a^x \dot{\varphi}(\xi)\, d\xi$.*

(a) $\int_J f\dot{\varphi}\, dm \le \sup|\varphi|\ \mathrm{Var}_J(f)$
(b) $\mathrm{Var}_J(f\varphi) \le \sup|\varphi|\ \mathrm{Var}_J(f) + \operatorname{ess\,sup}|\dot{\varphi}|\ \int_J |f|\, dm$
(c) $\int_J f\dot{\varphi}\, dm \le \sup_{u,v\in J}|\varphi(u) - \varphi(v)|\ \mathrm{var}_J(f) + \frac{\varphi(b)-\varphi(a)}{b-a}\int_J f\, dm.$
(d) If $\varphi(a) = \varphi(b) = 0$, *then* $\int_J f\dot{\varphi}\, dm \le \sup|\varphi|\ \mathrm{var}_J(f)$.
(e) $\mathrm{var}_J(f) \le \mathrm{Var}_J(f) \le 2\,\mathrm{var}_J(f) + \frac{2}{b-a}\left|\int_J f\, dm\right|$.

Before we prove this lemma, we discuss a number of consequences.

Corollary 2.1. $\int_J |f|\, dm \le \frac{1}{2}|J|\ \mathrm{Var}_J(f)$.

Proof. Let $\dot{\varphi} = 1_{\{f>0\}} - 1_{\{f<0\}}$. Then $\int_J |f|\, dm = \int_J f\dot{\varphi}\, dm \le \frac{1}{2}|J|\ \mathrm{Var}_J(f)$ by Lemma 2.2a applied to $\varphi(x) = -\int_a^{(a+b)/2}\dot{\varphi}(\xi)\, d\xi + \int_a^x \dot{\varphi}(\xi)\, d\xi$. □

Remark 2.2. It is easy to check that $\mathrm{Var}_I(.)$ and $\mathrm{var}_I(.)$ are seminorms, i.e. subadditive and positively homogeneous. Because of Corollary 2.1, $\|f\|_{BV} = \mathrm{Var}_I(f)$ defines indeed a norm on $BV(I) = \{f \in L^1_I : \mathrm{Var}_I(f) < \infty\}$. It is equivalent to the more common norm $\mathrm{var}_I(f) + \int_I |f|\, dm$ on $BV(I)$, see Lemma 2.2e.

Corollary 2.2. *Let $\varphi : J \to \mathbb{R}$ and suppose that the interval J is partitioned into subintervals $J_1, \ldots, J_r$ such that $\varphi|_{J_k}$ is continuously differentiable for each $k = 1, \ldots, r$. Let $f \in L^1_J$. Then*

$$\int_J f\varphi'\, dm \le 2\sup|\varphi|\left(\mathrm{var}_J(f) + \frac{1}{\min_k |J_k|}\int_J |f|\, dm\right)\ . \tag{24}$$

Proof. We apply Lemma 2.2c to each interval J_k separately:

$$\int_{J_k} f\varphi'\,dm \le 2\,\sup|\varphi|\,\left(\mathrm{var}_{J_k}(f)+\frac{1}{|J_k|}\int_{J_k}|f|\,dm\right)\,.$$

Summing over k and observing (19) this yields (24). □

Proof of Lemma 2.2. Let $\varepsilon>0$, and fix $s>1$ such that $\int_{\{|f|>s\}}|f|\,dm<\varepsilon$. There exists $\dot\psi\in C(J)$ such that $\int_J|\dot\varphi-\dot\psi|\,dm<\frac{\varepsilon}{s}$, $\sup|\dot\psi|\le \operatorname{ess\,sup}|\dot\varphi|+1$, and $\int_J\dot\varphi\,dm=\int_J\dot\psi\,dm$. Hence

$$\int_J f(\dot\varphi-\dot\psi)\,dm\le\varepsilon\operatorname{ess\,sup}|\dot\varphi-\dot\psi|+s\int_J|\dot\varphi-\dot\psi|\,dm\le C\varepsilon \tag{25}$$

where $C=2+2\operatorname{ess\,sup}|\dot\varphi|$.

Let $\psi(x)=c+\int_a^x\dot\psi(\xi)\,d\xi$. Then $\psi\in C^1(J)$ so that $\tilde\psi:=\psi/\sup|\psi|\in\mathcal{T}_J$. Hence $\int_J f\tilde\psi'\,dm\le\mathrm{Var}_J(f)$. As $\sup|\psi|\le\sup|\varphi|+\int_J|\dot\psi-\dot\varphi|\,dm\le\sup|\varphi|+\varepsilon$, it follows from (25) that

$$\int_J f\dot\varphi\,dm\le\int_J f\psi'\,dm+C\varepsilon\le(\sup|\varphi|+\varepsilon)\,\mathrm{Var}_J(f)+C\varepsilon\,.$$

As $\varepsilon>0$ is arbitrary, we conclude

$$\int_J f\dot\varphi\,dm\le\sup|\varphi|\;\mathrm{Var}_J(f) \tag{26}$$

which is part (a) of the lemma.

Now let $\psi\in C^1(J)$, $|\psi|\le1$. Then $(\varphi\psi)(x)=(\varphi\psi)(a)+\int_a^x(\dot\varphi\psi+\varphi\psi')(\xi)\,d\xi$, so we may write $\dot{(\varphi\psi)}=\dot\varphi\psi+\varphi\psi'$. Hence, in view of part (a),

$$\int_J(f\varphi)\psi'\,dm=\int_J f\,\dot{(\varphi\psi)}\,dm-\int_J f\dot\varphi\psi\,dm\le\sup|\varphi|\;\mathrm{Var}_J(f)+\operatorname{ess\,sup}|\dot\varphi|\int_J|f|\,dm\,.$$

This is part (b) of the lemma.

If $\varphi(a)=\varphi(b)=0$, then also $\psi(a)=\psi(b)=0$ so that $\psi\in\mathcal{T}_{J,0}$, and we can estimate by $\mathrm{var}_J(f)$ instead $\mathrm{Var}_J(f)$ on the right hand side of (26). This is part (d) of the lemma.

Next, for $\varphi\in\mathcal{T}_J$, let $\tilde\varphi(x)=\frac{x-a}{b-a}(\varphi(x)-\varphi(b))+\frac{b-x}{b-a}(\varphi(x)-\varphi(a))=\int_a^x(\varphi'(\xi)-\frac{\varphi(b)-\varphi(a)}{b-a})\,d\xi$. As $\tilde\varphi(a)=\tilde\varphi(b)=0$ we can apply part (d) of the lemma to $\tilde\varphi$. So

$$\int_J f\left(\varphi'-\frac{\varphi(b)-\varphi(a)}{b-a}\right)dm\le\sup|\tilde\varphi|\;\mathrm{var}_J(f)\le2\,\sup|\varphi|\;\mathrm{var}_J(f)\,. \tag{27}$$

As $\sup|\tilde\varphi|\le\sup_{u,v\in J}|\varphi(u)-\varphi(v)|$, this proves part (c) of the lemma. Finally, (27) also implies

$$\mathrm{Var}_J(f)=\sup_{\varphi\in\mathcal{T}_J}\int_J f\varphi'\,dm\le2\;\mathrm{var}_J(f)+\frac{2}{b-a}\left|\int_J f\,dm\right|$$

and $\mathrm{var}_J(f)\le\mathrm{Var}_J(f)$ follows directly from the definition. This proves part (e) of the lemma. □

2.4 The Lasota–Yorke Inequality

Coming back to the transfer operator P_τ we show next that $P_\tau(BV) \subseteq BV$. (Here and in the sequel we write BV instead of $BV(I)$ and $\mathrm{var}(f)$ instead of $\mathrm{var}_I(f)$.) In fact, we prove much more – an inequality which was discovered in this context by Lasota and Yorke [16]. In that paper, as well as in numerous subsequent generalizations of this result, the proof is based on the "elementary" approach (14) to variation. Here we give a proof using test functions as in (18).

Proposition 2.1 (Lasota–Yorke inequality). *Let $\tau : I \to I$ be a p.w.e. map as defined in Sect. 2.1. Let $\ell \in \mathbb{N}$ and recall that $\kappa_\ell := \inf |(\tau^\ell)'| > 0$ and $D_\ell = \sup |(\frac{1}{(\tau^\ell)'})'| < \infty$. Let also $E_\ell := 2/(\kappa_\ell \min_i |I_i^\ell|)$ where the I_i^ℓ's are monotonicity intervals of τ^ℓ which are finitely many. Then, for $f \in L^1_I$,*

$$\int_I |P_\tau f|\, dm \le \int_I |f|\, dm \tag{28}$$

$$\mathrm{Var}(P_\tau^\ell f) \le \frac{2}{\kappa_\ell}\, \mathrm{var}(f) + (D_\ell + E_\ell) \int_I |f|\, dm\,. \tag{29}$$

Proof. Equation (28) is just a restatement of (10). We turn to (29). As τ^ℓ is again a piecewise expanding map, it suffices to prove this estimate for $\ell = 1$. Let $\varphi \in \mathcal{T}_I$. As $(\varphi \circ \tau)'(x) = \varphi'(\tau(x))\, \tau'(x)$ for all $x \in I \setminus \{\zeta_0, \dots, \zeta_N\}$, we have

$$\begin{aligned} \int_I P_\tau f\, \varphi'\, dm &= \int_I f\, (\varphi' \circ \tau)\, dm = \int_I f\, \frac{(\varphi \circ \tau)'}{\tau'}\, dm \\ &= \int_I f\, \Big(\frac{\varphi \circ \tau}{\tau'}\Big)'\, dm - \int_I f\, (\varphi \circ \tau) \Big(\frac{1}{\tau'}\Big)'\, dm\,. \end{aligned} \tag{30}$$

The second term is bounded by $D_1 \int_I |f|\, dm$. To the first term we apply Corollary 2.2. As $|\tau'| \ge \kappa_1$, this yields (29) (for $\ell = 1$). □

If one applies inequality (29) to $P_\tau^\ell f, P_\tau^{2\ell} f, P_\tau^{3\ell} f, \dots$ and observes (28), one obtains by recurrence for each $k \in \mathbb{N}$

$$\mathrm{Var}(P_\tau^{k\ell} f) \le \Big(\frac{2}{\kappa_\ell}\Big)^k \mathrm{var}(f) + (D_\ell + E_\ell) \sum_{j=0}^{k-1} \Big(\frac{2}{\kappa_\ell}\Big)^j \int_I |f|\, dm\,. \tag{31}$$

As $\kappa_M > 2$ by assumption (see Sect. 2.1), it follows at once that

$$\mathrm{Var}(P_\tau^{kM} f) \le \Big(\frac{2}{\kappa_M}\Big)^k \mathrm{var}(f) + (D_M + E_M) \frac{\kappa_M}{\kappa_M - 2} \int_I |f|\, dm\,. \tag{32}$$

In order to extend this inequality to powers P_τ^n which are not multiples of M we decompose $n = kM + p$ with $0 \le p < M$. Equation (31) yields

$$\mathrm{Var}(P_\tau^p f) \le \left(\frac{2}{\kappa_1}\right)^p \mathrm{var}(f) + (D_1 + E_1)\frac{(2/\kappa_1)^M + 1}{|2/\kappa_1 - 1|}\int_I |f|\,dm \tag{33}$$

and combining this with (32) we arrive at

$$\|P_\tau^n f\|_{BV} = \mathrm{Var}(P_\tau^n f) \le C_1 \alpha^n \mathrm{var}(f) + C_2 \int_I |f|\,dm \tag{34}$$

where $0 < \alpha := (2/\kappa_M)^{\frac{1}{M}} < 1$ and C_1, C_2 are constants that depend on τ only through $M, \kappa_1, \kappa_M, D_1, D_M, E_1, E_M$. In particular, P_τ is a bounded linear operator on $(BV, \|.\|_{BV})$.

2.5 Compact Embedding and the Spectral Gap

The usefulness of the space BV is mainly due to the fact that it embeds compactly into L_I^1: the unit ball of BV is compact in L_I^1, that is, each sequence $(f_n)_n$ of L_I^1 functions with bounded BV-norm has a subsequence which converges (in L_I^1-norm) to an element of BV. It follows directly that $(BV, \|.\|_{BV})$ is complete, i.e. BV is a Banach space.

In its simplest form this is known as Helly's theorem. For the test function approach to variation that we follow here and that we will extend to multivariate functions in Chap. 3 this is proved e.g. in [13, 14, 15].[8]

A first consequence is the existence of an invariant probability density of bounded variation for the map τ: let $f_n := \frac{1}{n}\sum_{k=0}^{n-1} P_\tau^k 1$. Then $\|f_n\|_{BV} \le \frac{1}{n}\sum_{k=0}^{n-1} \|P_\tau^k 1\|_{BV}$, and (34) implies $\sup_n \|f_n\|_{BV} \le C_2 < \infty$. Hence there are $h \in L_m^1$ and a subsequence $(f_{n_j})_j$ such that $\lim_{j\to\infty} \int_I |h - f_{n_j}|\,dm = 0$. It follows from the elementary properties (9–11) of P_τ that h is a probability density and that the measure $\mu = hm$ is τ-invariant. The bound $\|h\|_{BV} \le C_2$ follows from (20).

But much more is true. The Lasota–Yorke inequality (29), together with the compact embedding property of BV into L_I^1, allows to apply the Ionescu-Tulcea/Marinescu theorem [17]:

Theorem 2.1 (Quasi-compactness of P_τ). *The operator $P_\tau : BV \to BV$ is* quasi-compact, *i.e. its canonical complexification has only finitely many eigenvalues of modulus one which all have finite multiplicities, and the rest of the spectrum is contained in a disc of radius $\rho < 1$ centered at 0. As seen before, 1 is an eigenvalue of P_τ. (We will fix ρ such that the rest of the spectrum is indeed contained in the interior of the disc of radius ρ.)*[9]

[8] It is a simple exercise to derive the compact embedding of BV into L_I^1 from Lemma 2.2. *Hint:* Subdivide I into 2^n intervals of length 2^{-n}. For $f \in BV$ let $f_n = \sum_{k=1}^{2^n} 1_{I_k} 2^n \int_{I_k} f\,dm$. Let $\dot\varphi_n = \mathrm{sign}(f - f_n)$. Then $\int_I |f - f_n|\,dm \le \sum_{k=1}^{2^n} 2^{-n}\,\mathrm{var}_{I_k}(f - f_n) \le 2^{-n}\,\mathrm{var}_I(f)$ by Lemma 2.2c and (18).

[9] More detailed accounts of this theorem can be found e.g. in [11, Chap. 7] and [18, Sect. 3.2]. See also [19].

From now on we assume that τ is *mixing*. That means that 1 is a simple eigenvalue of P_τ and that there is no other eigenvalue of modulus one. For p.w.e. maps this in fact equivalent to the usual notion of mixing in ergodic theory, see e.g. [11, Corollary 7.2.1]. For mixing τ, $(1-\rho)$ quantifies a *spectral gap*, i.e. the simple eigenvalue 1 is separated from the moduli of all other spectral value by $(1-\rho)$ at least. We have indeed

Corollary 2.3 (Spectral gap of P_τ). *If τ is mixing, then there is a constant $C_3 > 0$ such that*

$$\int_I |P_\tau^n f|\, dm \le \|P_\tau^n f\|_{BV} \le C_3\, \rho^n\, \|f\|_{BV} \tag{35}$$

for all $n \in \mathbb{N}$ and all $f \in BV$ with $\int_I f\, dm = 0$.[10]

Remark 2.3. Although we will not use it explicitly we note the following fact: both constants ρ and C_3 do not change much under small perturbations of τ as long as τ and its perturbations satisfy a Lasota–Yorke inequality (34) with the same constants α, C_1, and C_2; see [20].)

3 Finite Systems

As an intermediate step towards infinite coupled systems, this section deals with finite coupled systems of d piecewise expanding maps described by a transformation T_ϵ on the d-dimensional unit cube. We will see below that – for sufficiently small $|\epsilon|$ – the maps T_ϵ are piecewise expanding and that one can develop a spectral theory for their transfer operators P_{T_ϵ} in just the same way as we did it for the 1D map τ in Chap. 2.

3.1 The Coupling

We recall the notation from the Introduction:

- L is a finite set of cardinality $d > 0$: it serves as the set of sites. For notational convenience we work with $L = \{1, \dots, d\}$ in this section without interpreting L as a subset of the one-dimensional lattice $\mathbb{Z}$.
- $X = I^L$ is the state space of the system: it is a d-dimensional cube.
- $\tau : I \to I$ is a p.w.e. map as defined in Sect. 2.1.

[10] In the spectral theoretic approach the constants C_3 and ρ cannot be determined easily from the "formula" for the map τ. For some maps explicit estimates for ρ with $C_3 = 1$ are derived in [21]. The proof, which is a refined version of the proof of our Lemma 2.2c, bypasses spectral theory completely. In [22] (see also [23, Sect. 8]) it is shown how to obtain explicit estimates on C_3 and ρ using Birkhoff cones. A rigorous numerical approach to estimate these constants is discussed in [23].

- $T_0 : X \to X$ is the d-fold product of the map τ: $(T_0 \boldsymbol{x})_i = \tau(x_i)$ $(i \in L)$.
- $\Phi_\epsilon : X \to X$ $(|\epsilon| < \epsilon_0)$ is a family of coupling maps ϵ-close to the identity in a C^2 sense made precise below.
- $T_\epsilon = \Phi_\epsilon \circ T_0 : X \to X$ are the maps describing the coupled systems.

The precise assumptions on Φ_ϵ are: $\Phi_\epsilon(\boldsymbol{x}) = \boldsymbol{x} + A_\epsilon(\boldsymbol{x})$ is a (a_1, a_2)-*coupling*, i.e. there are $L \times L$ matrices A', A'' with $a_1 = \|A'\|_1$, $a_2 = \|A''\|_1$ (maximal column sum norm) such that for all $i, j, k \in L$

$$|(A_\epsilon)_i| \le 2|\epsilon|, \quad |(DA_\epsilon)_{ij}| \le 2|\epsilon| A'_{ij}, \quad |\partial_k (DA_\epsilon)_{ij}| \le 2|\epsilon| A''_{ij} \,. \tag{36}$$

Here ∂_j denotes the partial derivative w.r.t. x_j. The diffusive nearest neighbor coupling (3) is an example of a $(1, 0)$-coupling.

Later we will need the following estimates on $(D\Phi_\epsilon)^{-1}$ derived from (36):

$$|((D\Phi_\epsilon)^{-1})_{ij}| \le ((E - 2|\epsilon|A')^{-1})_{ij} \quad \text{where } E \text{ is the identity matrix,} \tag{37}$$

$$\sum_{i=1}^{d} |((D\Phi_\epsilon)^{-1})_{ij}| \le \frac{1}{1 - 2a_1|\epsilon|} \,, \quad \sum_{i=1}^{d} |\partial_i ((D\Phi_\epsilon)^{-1})_{ij}| \le \frac{2a_2|\epsilon|}{(1 - 2a_1|\epsilon|)^2} \,. \tag{38}$$

Observe first that $(D\Phi_\epsilon)^{-1} = \sum_{n=0}^{\infty} (-DA_\epsilon)^n$ and that $(-DA_\epsilon)^n$ is dominated coefficient-wise by $|2\epsilon|^n A'^n$ in view of (36). This yields (37). Let $\mathbf{1} = (1, \dots, 1)$ and let $\boldsymbol{e}_j$ be the j-th unit vector. We interpret both as *matrices*, which plays a role when we evaluate their $\|.\|_1$-norms. Then

$$\sum_{i=1}^{d} |((D\Phi_\epsilon)^{-1})_{ij}| \le \sum_{n=0}^{\infty} |2\epsilon|^n \mathbf{1} A'^n \boldsymbol{e}_j \le \sum_{n=0}^{\infty} |2\epsilon|^n \|\mathbf{1}\|_1 \|A'\|_1^n \|\boldsymbol{e}_j\|_1 = \frac{1}{1 - 2a_1|\epsilon|} \,.$$

This is the first estimate in (38), and the second one is proved along the same lines.

3.2 The Transfer Operator

Recall from Sect. 2.1 that we denote the intervals restricted to which the map τ is C^2 by $I_1, \dots, I_N$. Let $\mathcal{Q}_d = \{I_{i_1} \times \cdots \times I_{i_d} : \; i_1, \dots, i_d \in \{1, \dots, N\}\}$ be the family of rectangular domains restricted to which the product map T_0 is C^2. As in (7) we define the transfer operator P_{T_0} of T_0 acting on measurable $f : X \to \mathbb{R}$ by

$$P_{T_0} f(\boldsymbol{x}) = \sum_{Q \in \mathcal{Q}_d} \frac{f}{|\det(DT_0)|} \circ (T_0|_Q)^{-1}(\boldsymbol{x}) \cdot 1_{T_0(Q)}(\boldsymbol{x}) \,. \tag{39}$$

As in the one-dimensional case, P_{T_0} can be interpreted as a positive linear contraction on the space L^1_X of equivalence classes of Lebesgue integrable functions from X to $\mathbb{R}$, unambiguously defined by

$$\int_X P_{T_0} f(\boldsymbol{x})\, \psi(\boldsymbol{x})\, d\boldsymbol{x} = \int_X f(\boldsymbol{x})\, \psi(T_0(\boldsymbol{x}))\, d\boldsymbol{x} \tag{40}$$

where $d\boldsymbol{x}$ is an abbreviation for $dx_1 \dots dx_d$.

In the same way we define a transfer operator P_{Φ_ϵ} for the coupling map Φ_ϵ. Since Φ_ϵ is injective, its explicit form is particularly simple:

$$P_{\Phi_\epsilon} f(\boldsymbol{x}) = \frac{f}{|\det(D\Phi_\epsilon)|} \circ \Phi_\epsilon^{-1}(\boldsymbol{x}) \cdot 1_{\Phi_\epsilon(X)}(\boldsymbol{x})\ . \tag{41}$$

In view of the elementary properties of general transfer operators discussed in Remark 2.1, we have for the transfer operator P_{T_ϵ} of the coupled map $T_\epsilon = \Phi_\epsilon \circ T_0$

$$P_{T_\epsilon} = P_{\Phi_\epsilon} P_{T_0} \tag{42}$$

and both, P_{Φ_ϵ} and P_{T_ϵ}, have a (pre)-dual characterization as linear L^1_X operators analogous to (40).

3.3 Multivariate Functions of Bounded Variation

As in the 1D case we need a subspace of L^1_X of more regular functions on which the transfer operators just introduced have "good" spectral properties. Multivariate functions of bounded variation turn out to be a suitable choice.

There are many equivalent ways to define the variation of a multivariate function $f : X \to \mathbb{R}$, see e.g. [13, 14, 15]. The most intuitive one is perhaps to define it just in terms of coordinate-wise one-dimensional variation of f. To this end, and also for later use, we introduce the following notation: For $i \in \{1, \dots, d\}$ we identify $\boldsymbol{x}$ and $(x_i, \boldsymbol{x}_{\neq i})$ where $\boldsymbol{x}_{\neq i} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_d)$. Since we never permute coordinates, this will not lead to any confusion. We also denote by $X_{\neq i}$ the $(d-1)$-dimensional cube $\{\boldsymbol{x}_{\neq i} : \boldsymbol{x} \in X\}$ and by $f_{\boldsymbol{x}_{\neq i}} : I \to \mathbb{R}$, $f_{\boldsymbol{x}_{\neq i}}(x) = f(x, \boldsymbol{x}_{\neq i})$, the $\boldsymbol{x}_{\neq i}$-section of f. Now we define for $f \in L^1_X$

$$\mathrm{Var}^i_X(f) = \int_{X_{\neq i}} \mathrm{Var}_I(f_{\boldsymbol{x}_{\neq i}})\, d\boldsymbol{x}_{\neq i}, \quad \mathrm{Var}_X(f) = \max_{i=1,\dots,d} \mathrm{Var}^i_X(f)\ . \tag{43}$$

Observe that $f_{\boldsymbol{x}_{\neq i}} \in L^1_I$ for Lebesgue-a.e. $\boldsymbol{x}_{\neq i} \in X_{\neq i}$ by Fubini's theorem. So $\mathrm{Var}(f_{\boldsymbol{x}_{\neq i}}) \in [0, \infty]$ is well defined for Lebesgue-a.e. $\boldsymbol{x}_{\neq i}$. That it depends measurably on $\boldsymbol{x}_{\neq i}$ will be shown in Lemma 3.1b. We note the following immediate consequences of Lemma 2.2e and Corollary 2.1:

$$\mathrm{Var}^i_X(1) = 2 \quad \text{and} \quad \int_X |f|\, dm \le \frac{1}{2} \mathrm{Var}^i_X(f) \quad \text{for each } i = 1, \dots, d, \tag{44}$$

where m denotes Lebesgue measure on X.

Let

$$BV(X) = \{f \in L^1_X : \mathrm{Var}_X(f) < \infty\} \tag{45}$$

be the space of functions of bounded variation on X, and note that $\mathrm{Var}_X(.)$ is a norm on $BV(X)$.

The above definition of $\mathrm{Var}^i_X(f)$ is equivalent to a more direct one generalizing the test function approach (18) we used already in dimension one. Let

$$\mathcal{T}_X = \{\varphi \in C^1(X) : |\varphi| \le 1\} . \tag{46}$$

Proposition 3.1. *For each measurable $f : X \to \mathbb{R}$ and each $i = 1, \dots, d$,*

$$\mathrm{Var}^i_X(f) = \sup_{\varphi \in \mathcal{T}_X} \int_X f(\boldsymbol{x})\, \partial_i \varphi(\boldsymbol{x})\, d\boldsymbol{x} . \tag{47}$$

An immediate consequence is that

$$\mathrm{Var}_X(f) \le \liminf_{n\to\infty} \mathrm{Var}_X(f_n) \tag{48}$$

whenever $f, f_n \in L^1_X$ and $\lim_{n\to\infty} \int |f - f_n|\, dm = 0$.

The proof of this proposition requires smoothing of functions and test functions by *mollifiers*: Let $\eta : \mathbb{R} \to [0, \infty)$ be a symmetric (at zero) C^∞ function with $\int_{\mathbb{R}} \eta(t)\, dt = 1$ and $\eta(t) = 0$ if $|t| \ge 1$. For $\delta > 0$ let $\eta_\delta(t) = \delta^{-1}\, \eta(\frac{t}{\delta})$. The *convolution* of a function $u : I \to \mathbb{R}$ with η_δ is defined by $(u * \eta_\delta)(x) = \int_{\mathbb{R}} u(x - t)\eta_\delta(t)\, dt$, where $u(x - t)$ is understood to be zero if $(x - t) \notin I$.

Lemma 3.1. *Let $f_{\boldsymbol{x}_{\neq i},\delta} = f_{\boldsymbol{x}_{\neq i}} * \eta_\delta$.*

(a) $\mathrm{Var}_I(f_{\boldsymbol{x}_{\neq i}}) = \lim_{\delta\to 0} \mathrm{Var}_I(f_{\boldsymbol{x}_{\neq i},\delta})$ *for every* $\boldsymbol{x}_{\neq i} \in X_{\neq i}$.
(b) $\boldsymbol{x}_{\neq i} \mapsto \mathrm{Var}_I(f_{\boldsymbol{x}_{\neq i},\delta})$ *and* $\boldsymbol{x}_{\neq i} \mapsto \mathrm{Var}_I(f_{\boldsymbol{x}_{\neq i}})$ *are nonnegative measurable functions. In particular,* $\mathrm{Var}^i_X(f)$ *is well defined in (43).*
(c) $\int_{X_{\neq i}} \mathrm{Var}_I(f_{\boldsymbol{x}_{\neq i},\delta})\, d\boldsymbol{x}_{\neq i} \le \sup_{\varphi\in\mathcal{T}_X} \int_X f(\boldsymbol{x})\partial_i\varphi(\boldsymbol{x})\, d\boldsymbol{x} + o(\delta)$.

Proof. (a) It is a rather classical result from real analysis [14, Theorem 1.6.1] that for each $\boldsymbol{x}_{\neq i} \in X_{\neq i}$

$$\lim_{\delta\to 0} \int_I |f_{\boldsymbol{x}_{\neq i},\delta}(x_i) - f_{\boldsymbol{x}_{\neq i}}(x_i)|\, dx_i = 0 . \tag{49}$$

This implies at once that

$$\mathrm{Var}_I(f_{\boldsymbol{x}_{\neq i}}) \le \liminf_{\delta\to 0} \mathrm{Var}_I(f_{\boldsymbol{x}_{\neq i},\delta}) . \tag{50}$$

We turn to the reverse inequality. Let $\varphi \in \mathcal{T}_I$, $\varepsilon > 0$, and let $\tilde\varphi$ be any C^1 extension of φ to all of $\mathbb{R}$ with $|\tilde\varphi| \le 1 + \varepsilon$. Then $\tilde\varphi * \eta_\delta|_I \in (1 + \varepsilon)\, \mathcal{T}_I$. So

$$\begin{aligned}\int_I f_{\boldsymbol{x}_{\neq i},\delta}(x_i)\varphi'(x_i)\,dx_i &= \int_I (f_{\boldsymbol{x}_{\neq i}} * \eta_\delta)(x_i)\tilde{\varphi}'(x_i)\,dx_i \\ &= \int_I f_{\boldsymbol{x}_{\neq i}}(x_i)(\tilde{\varphi}' * \eta_\delta)(x_i)\,dx_i \\ &= \int_I f_{\boldsymbol{x}_{\neq i}}(x_i)(\tilde{\varphi} * \eta_\delta)'(x_i)\,dx_i \le \mathrm{Var}_I(f_{\boldsymbol{x}_{\neq i}}) \quad (51)\end{aligned}$$

by definition in (23). Hence $\mathrm{Var}_I(f_{\boldsymbol{x}_{\neq i},\delta}) \le \mathrm{Var}_I(f_{\boldsymbol{x}_{\neq i}})$.

(b) $\mathrm{Var}_I(f_{\boldsymbol{x}_{\neq i},\delta}) = \int_I |f'_{\boldsymbol{x}_{\neq i},\delta}|\,dm$ is a nonnegative measurable function of the argument $\boldsymbol{x}_{\neq i}$, and so is $\mathrm{Var}_I(f_{\boldsymbol{x}_{\neq i}})$ in view of part (a) of this lemma.

(c) Let $\psi_{\boldsymbol{x}_{\neq i}}(x_i) = \mathrm{sign}(f'_{\boldsymbol{x}_{\neq i},\delta}(x_i))$, and let $J = [-1, 2]$. By Lemma 2.1,

$$\mathrm{Var}_I(f_{\boldsymbol{x}_{\neq i},\delta}) \le \int_J |f'_{\boldsymbol{x}_{\neq i},\delta}(x_i)|\,dx_i = \int_J (f_{\boldsymbol{x}_{\neq i}} * \eta_\delta)'(x_i)\,\psi_{\boldsymbol{x}_{\neq i}}(x_i)\,dx_i\,. \quad (52)$$

Let $\varepsilon > 0$. As $(f_{\boldsymbol{x}_{\neq i}} * \eta_\delta)'$ is bounded and as $|\psi_{\boldsymbol{x}_{\neq i}}| \le 1$, there is $\varphi \in C^1(X_{\neq i} \times J)$ with $|\varphi| \le 1$ such that

$$\begin{aligned}&\int_{X_{\neq i}} \int_J (f_{\boldsymbol{x}_{\neq i}} * \eta_\delta)'(x_i)\,\psi_{\boldsymbol{x}_{\neq i}}(x_i)\,dx_i d\boldsymbol{x}_{\neq i} \\ &\quad \le \varepsilon + \int_{X_{\neq i}} \int_J (f_{\boldsymbol{x}_{\neq i}} * \eta_\delta)'(x_i)\,\varphi_{\boldsymbol{x}_{\neq i}}(x_i)\,dx_i d\boldsymbol{x}_{\neq i} \quad (53) \\ &\quad = \varepsilon + \int_{X_{\neq i}} \int_J f_{\boldsymbol{x}_{\neq i}}(x_i)\,(\varphi_{\boldsymbol{x}_{\neq i}} * \eta_\delta)'(x_i)\,dx_i d\boldsymbol{x}_{\neq i}\,.\end{aligned}$$

For the last identity observe that $(f_{\boldsymbol{x}_{\neq i}} * \eta_\delta)(x_i) = 0$ if $x_i \notin (-\delta, 1+\delta)$. As $f_{\boldsymbol{x}_{\neq i}}(x_i) = 0$ for $x_i \notin I$, the integral over J in the last expression can be replaced by an integral over I. Define $\tilde{\varphi} : X \to \mathbb{R}$ by $\tilde{\varphi}(\boldsymbol{x}) = (\varphi_{\boldsymbol{x}_{\neq i}} * \eta_\delta)(x_i)$. Clearly, $\varphi \in \mathcal{T}_X$, and combining (52) with (53) we obtain

$$\int_{X_{\neq i}} \mathrm{Var}_I(f_{\boldsymbol{x}_{\neq i},\delta})\,d\boldsymbol{x}_{\neq i} \le \varepsilon + \int_X f(\boldsymbol{x})\partial_i\tilde{\varphi}(\boldsymbol{x})\,d\boldsymbol{x}\,.$$

As $\varepsilon > 0$ is arbitrary, this finishes the proof. □

Proof of Proposition 3.1. Fix f and i and denote the expression on the right hand side of (47) by $V(f)$. Then $V(f) \le \mathrm{var}_X^i(f)$ because, for $\varphi \in \mathcal{T}_X$, all $\varphi_{\boldsymbol{x}_{\neq i}}$ $(\boldsymbol{x}_{\neq i} \in X_{\neq i})$ belong to the set $\mathcal{T}_I$ of univariate test functions, see (22).

The reverse inequality follows at once from Lemma 3.1a, Fatou's lemma, and Lemma 3.1c. □

3.4 The Lasota–Yorke Inequality

The Lasota–Yorke inequality (29) for iterates of one-dimensional p.w.e. maps involves three constants on its right hand side which are determined by basic properties of the map τ^ℓ:

- κ_ℓ is the minimal slope of τ^ℓ, i.e. κ_ℓ^{-1} is an upper bound on the contraction rate of each single inverse branch of τ^ℓ,
- D_ℓ is determined essentially by the second derivative of τ, and
- E_ℓ can be controlled in terms of the inverse of the minimal size of intervals of monotonicity of τ^ℓ.[11]

The following proposition shows that these are also the quantities one needs to control for a Lasota–Yorke type inequality for $P^\ell_{T_\epsilon}$.

Proposition 3.2 (Lasota–Yorke inequality for finite coupled systems). *Let $\ell \in \mathbb{N}$. For each $\alpha_\ell > \frac{2}{\kappa_\ell}$ and each $C_{4,\ell} > D_\ell + E_\ell$ there is $\epsilon_1 \in (0, \epsilon_0]$ such that for $|\epsilon| \le \epsilon_1$*

$$\int_X |P_{T_\epsilon} f|\, dm \le \int_X |f|\, dm \tag{54}$$

$$\operatorname{Var}_X(P^\ell_{T_\epsilon} f) \le \alpha_\ell \operatorname{Var}_X(f) + C_{4,\ell} \int_X |f|\, dm\,. \tag{55}$$

Given ℓ, α_ℓ and $C_{4,\ell}$, the choice of ϵ_1 depends only on the constants a_1, a_2 which qualify Φ_ϵ as a (a_1, a_2)-coupling, see (36).

Equation (54) follows again from (10), see also the remark thereafter. For (55) we will give a complete proof only when $\ell = 1$. In this case it follows directly from the following separate estimates for P_{T_0} and P_{Φ_ϵ}.

Lemma 3.2.

$$\operatorname{Var}_X(P^\ell_{T_0} f) \le \frac{2}{\kappa_\ell} \operatorname{Var}_X(f) + (D_\ell + E_\ell) \int_X |f|\, dm\,. \tag{56}$$

Lemma 3.3.

$$\operatorname{Var}_X(P_{\Phi_\epsilon} f) \le \frac{1}{1 - 2a_1|\epsilon|} \operatorname{Var}_X(f) + \frac{2a_2|\epsilon|}{(1 - 2a_1|\epsilon|)^2} \int_X |f|\, dm\,. \tag{57}$$

Remark 3.1. As the Lasota–Yorke inequality in Proposition 3.2 is useful only if $\kappa_\ell = \inf |(\tau^\ell)'| > 2$, the restriction to the case $\ell = 1$ means that we assume $\inf |\tau'| > 2$. This was the case dealt with in [6]. It was only in the unpublished thesis [7] that the geometrically much more subtle case of general ℓ was dealt with.[12]

[11] A closer look at the definition of E_ℓ in Proposition 2.1 reveals that one can do better: It is essentially the minimal size of the *images* of the intervals of monotonicity which determines E_ℓ.

[12] The treatment of this case in [7] is based on an alternative proof of Lemma 3.2 as given in [6, Lemma 3.1]. Instead of using the product structure of T_0^ℓ it suffices to use the fact that the domains restricted to which T_0^ℓ is C^2 and expanding are direct products of intervals (on which τ^ℓ is monotone and C^2). Let us call this the *rectangular domain property*. (The proof of [6, Lemma 3.1] is rather

Proof of Lemma 3.2. The map $T_0 = \tau \times \cdots \times \tau$ is the d-fold direct product of the p.w.e. map τ, so $T_0^\ell = \tau^\ell \times \cdots \times \tau^\ell$ is the d-fold direct product of the p.w.e. map τ^ℓ. Therefore, without loss of generality, we may just treat the case $\ell = 1$ in this lemma.

As $\mathrm{Var}_X(f) = \max_{i=1,\dots,d} \mathrm{Var}_X^i(f)$, it suffices to prove inequality (56) for each Var_X^i separately. We will make as complete use as possible of the product structure of T_0. For notational simplicity we will estimate $\mathrm{Var}_X^1(P_{T_0} f)$ only, the other $\mathrm{Var}_X^i(P_{T_0} f)$ are treated in just the same way.

We write T_0 as $T_0 = S_2 \circ S_1$ where $S_1 = \mathrm{Id}_I \times (\tau \times \cdots \times \tau)$ and $S_2 = \tau \times (\mathrm{Id}_I \times \cdots \times \mathrm{Id}_I)$. Then $P_{T_0} = P_{S_2} P_{S_1}$, and we can do the estimate in two steps.

We start by estimating $\mathrm{Var}_X^1(P_{S_2} f)$ for $f \in L_X^1$. Because of the product structure of S_2, the operator P_{S_2} acts formally like a tensor product operator on L_X^1. More precisely, $(P_{S_2} f)_{\boldsymbol{x}_{\neq 1}}(x_1) = (P_\tau f_{\boldsymbol{x}_{\neq 1}})(x_1)$. Hence

$$\begin{aligned}
\mathrm{Var}_X^1(P_{S_2} f) &= \int_{X_{\neq 1}} \mathrm{Var}_I((P_{S_2} f)_{\boldsymbol{x}_{\neq 1}})\, d\boldsymbol{x}_{\neq 1} = \int_{X_{\neq 1}} \mathrm{Var}_I(P_\tau f_{\boldsymbol{x}_{\neq 1}})\, d\boldsymbol{x}_{\neq 1} \\
&\le \frac{2}{\kappa_1} \int_{X_{\neq 1}} \mathrm{Var}_I(f_{\boldsymbol{x}_{\neq 1}})\, d\boldsymbol{x}_{\neq 1} + (D_1 + E_1) \int_{X_{\neq 1}} \int_I |f_{\boldsymbol{x}_{\neq 1}}(x_1)|\, dx_1 d\boldsymbol{x}_{\neq 1} \\
&= \frac{2}{\kappa_1} \mathrm{Var}_X^1(f) + (D_1 + E_1) \int_X |f|\, dm\,.
\end{aligned}$$

Here we used the Lasota–Yorke inequality (29) for 1D maps from Proposition 2.1. Hence,

$$\mathrm{Var}_X^1(P_{T_0} f) = \mathrm{Var}_X^1(P_{S_2}(P_{S_1} f)) \le \frac{2}{\kappa_1} \mathrm{Var}_X^1(P_{S_1} f) + (D_1 + E_1) \int_X |P_{S_1} f|\, dm\,.$$

As $\int_X |P_{S_1} f|\, dm \le \int_X |f|\, dm$ (compare (10)), the proof of Lemma 3.2 will be finished by showing that $\mathrm{Var}_X^1(P_{S_1} f) \le \mathrm{Var}_X^1(f)$: let $\varphi \in \mathcal{T}_X$. For $\boldsymbol{x}_{\neq 1} \in X_{\neq 1}$ let $\psi_{\boldsymbol{x}_{\neq 1}}(x_1) = \varphi(S_1(x_1, \boldsymbol{x}_{\neq 1}))$. Then $\psi_{\boldsymbol{x}_{\neq 1}} \in \mathcal{T}_I$, and

straightforward analysis.) Now, if one passes to coupled systems, things change. Although $T_\epsilon = \Phi_\epsilon \circ T_0$ still possesses the rectangular domain property, this is no longer true for powers T_ϵ^ℓ, $\ell \ge 2$. In fact, already to make sure that, by passing from T_0 to T_ϵ, no new domains occur one needs the full strength of the regularity assumption (1). But with this assumption one can prove a geometrically much finer result [7]:

There are constants $\tilde{\epsilon} > 0$ and $c > 0$, *independent of the size of* L, such that for $|\epsilon| \le \tilde{\epsilon}$ and for each domain Z_ϵ on which T_ϵ^M is C^2 and expanding, there is a diffeomorphism Ψ_{Z_ϵ} between Z_ϵ and the corresponding rectangular domain Z_0 of T_0^M which is C^2 close to the identity in the sense of a $(1, c)$-coupling, see (36). This allows to reduce variation estimates of functions $f\, 1_{Z_\epsilon}$ to variation estimates of functions $\tilde{f}\, 1_{Z_0}$, and the latter ones can be dealt with using [6, Lemma 3.1]. See also [24] for a more details.

$$\int_X P_{S_1} f(\boldsymbol{x}) \partial_1 \varphi(\boldsymbol{x})\, d\boldsymbol{x} = \int_X f(\boldsymbol{x}) \partial_1 \varphi(S_1(\boldsymbol{x}))\, d\boldsymbol{x}$$
$$= \int_{X_{\neq 1}} \int_I f_{\boldsymbol{x}_{\neq 1}}(x_1)\, \psi'_{\boldsymbol{x}_{\neq 1}}(x_1)\, dx_1 d\boldsymbol{x}_{\neq 1} \le \int_{X_{\neq 1}} \mathrm{Var}_I(f_{\boldsymbol{x}_{\neq 1}})\, d\boldsymbol{x}_{\neq 1} = \mathrm{Var}^1_X(f)\ .$$

Now $\mathrm{Var}^1_X(P_{S_1} f) \le \mathrm{Var}^1_X(f)$ follows from Proposition 3.1. □

Proof of Lemma 3.3. Let $f \in L^1_X$, $\varphi \in \mathcal{T}_X$, and $j \in \{1, \dots, d\}$. Denote (just for this proof) the matrix $(D\Phi_\epsilon)^{-1}$ by (b_{ij}). Then

$$\begin{aligned}(\partial_j \varphi) \circ \Phi_\epsilon = (D(\varphi \circ \Phi_\epsilon)(D\Phi_\epsilon)^{-1})_j &= \sum_{i=1}^d \partial_i(\varphi \circ \Phi_\epsilon) \cdot b_{ij} \\ &= \sum_{i=1}^d \partial_i(\varphi \circ \Phi_\epsilon \cdot b_{ij}) - \sum_{i=1}^d \varphi \circ \Phi_\epsilon \cdot \partial_i b_{ij}\ .\end{aligned} \tag{58}$$

Let $\psi_{ij} = \varphi \circ \Phi_\epsilon \cdot b_{ij} = \varphi \circ \Phi_\epsilon \cdot ((D\Phi_\epsilon)^{-1})_{ij}$. As all functions ψ_{ij} are in $C^1(X)$, this implies

$$\begin{aligned}\int_X P_{\Phi_\epsilon} f(\boldsymbol{x})\, \partial_j \varphi(\boldsymbol{x})\, d\boldsymbol{x} &= \int_X f(\boldsymbol{x})\, \partial_j \varphi(\Phi_\epsilon \boldsymbol{x})\, d\boldsymbol{x} \\ &\le \sum_{i=1}^d \sup_{\boldsymbol{x}} |\psi_{ij}(\boldsymbol{x})|\, \mathrm{Var}_X(f) + \sum_{i=1}^d \sup_{\boldsymbol{x}} |\partial_i b_{ij}(\boldsymbol{x})| \int_X |f|\, dm\ .\end{aligned} \tag{59}$$

The two suprema in this estimate are precisely controlled by our assumptions (36) on Φ_ϵ and their consequences (38): for $j = 1, \dots, d$,

$$\begin{aligned}\sum_{i=1}^d \sup_{\boldsymbol{x}} |\psi_{ij}(\boldsymbol{x})| \le \sum_{i=1}^d \sup_{\boldsymbol{x}} |((D\Phi_\epsilon)^{-1})_{ij}(\boldsymbol{x})| &\le \frac{1}{1 - 2a_1|\epsilon|} \\ \sum_{i=1}^d \sup_{\boldsymbol{x}} |\partial_i b_{ij}(\boldsymbol{x})| = \sum_{i=1}^d \sup_{\boldsymbol{x}} |\partial_i((D\Phi_\epsilon)^{-1})_{ij}(\boldsymbol{x})| &\le \frac{2a_2|\epsilon|}{(1 - 2a_1|\epsilon|)^2}\ .\end{aligned}$$

As φ is an arbitrary test function in $\mathcal{T}_X$, this finishes the proof of Lemma 3.3. □

3.5 Existence of Absolutely Continuous Invariant Measures

Having derived inequality (55) one can proceed as in the one-dimensional case: as in Sect. 2.4 it follows that

$$\mathrm{Var}_X(P^n_{T_\epsilon} f) \le 2C_1\, \alpha^n\, \mathrm{Var}_X(f) + 2C_2 \int_X |f|\, dm \tag{60}$$

for all $f \in L^1_X$ and all $n \in \mathbb{N}$ provided $|\epsilon| \le \epsilon_1$. The constants are from (34), and the additional factor 2 accounts for the passage from $\epsilon = 0$ to $|\epsilon| \le \epsilon_1$.

(Indeed, there is nothing special with the factor 2. Any factor strictly larger than 1 would do as well.)

In perfect analogy with the case of piecewise expanding maps of the interval Proposition 3.2 immediately implies the existence of at least one absolutely continuous invariant measure.

Theorem 3.1. *Let $T_\epsilon : X \to X$ be a coupled map on $X = I^L$ as described in Sect. 3.1. For each $|\epsilon| \le \epsilon_1$ there exists a T_ϵ-invariant probability measure μ_ϵ which belongs to $BV(X)$.*

Proof. As $Var_X(1) = 2$ inequality (60) implies $\limsup_{n\to\infty} \mathrm{Var}_X(P^\ell_{T_\epsilon} 1) \le 2C_2$. Let $h_n = \frac{1}{n}\sum_{k=0}^{n-1} P^k_{T_\epsilon} 1$. Then also $\limsup_{n\to\infty} \mathrm{Var}_X(P^k_{T_\epsilon} 1) \le 2C_2$. As the space $BV(X)$ embeds compactly into L^1_X, see e.g. [13, Theorem 1.19] or [14, Corollary 5.3.4], it follows that $\{h_n\}$ has accumulation points in L^1 which belong to BV and are the density of an invariant measure. □

Although we will not use this observation explicitly, an argument of the same type will guarantee the existence of an invariant measure with marginal densities of bounded variation for the infinite coupled system in Chap. 4.

For sufficiently small $|\epsilon|$ the operator P_{T_ϵ} is again quasi-compact on the Banach space $BV(X)$, and one can show that it has a spectral gap if the single site map τ is mixing.[13] It is not possible, however, to obtain in this way a useful d-dependent control over the constants C_3 and ρ in the spectral gap estimate (35). To achieve this we will apply a more recent technique in Chap. 4. As a by-product we obtain the following d-dependent estimate on the mixing rate for uncoupled systems: for $f \in BV(X)$ with $\int_X f\, dm = 0$,

$$\int_X |P^n_{T_0} f|\, dm \le (2 + C_2) C_3\, d\, \rho^n\ \mathrm{Var}_X(f) \tag{61}$$

with constants C_2, C_3, and ρ from (34) and Corollary 2.3. This is proved at the end of Sect. 4.5.

4 Infinite Systems over $L = \mathbb{Z}$

The first problem that comes to mind if one attempts to transfer the finite system theory from Chap. 3 to the case $L = \mathbb{Z}$ is certainly: what is a class of measures which can play the role that the absolutely continuous ones play in the finite-dimensional case? These are not the measures absolutely continuous w.r.t. the infinite product Lebesgue measure $m^{\mathbb{Z}}$ on the "infinite-dimensional unit cube" $X = I^{\mathbb{Z}}$. Just look at the uncoupled map T_0: If $\mu = hm$ is an invariant measure for the p.w.e. map τ, then its infinite product $\mu^{\mathbb{Z}}$ should be

[13] For the case when τ is a mixing tent map a proof is published in [25].

the measure to look at. But $\mu^{\mathbb{Z}}$ is absolutely continuous w.r.t. $m^{\mathbb{Z}}$ if and only if μ is the Lebesgue measure on I, i.e. if $h = 1$.[14]

So we are lead to look at measures whose finite-dimensional marginals are absolutely continuous. We will introduce various norms on spaces of such measures, derive Lasota–Yorke inequalities for the transfer operator of T_ϵ on such spaces and prove the existence of spectral gaps.

4.1 Classes of Measures and Distributions

Let us fix some notation:

- $X = I^{\mathbb{Z}}$, and $\mathcal{M}$ is the space of signed Borel measures on X.
- For $\Lambda' \subset \Lambda \subset \mathbb{Z}$, let $\pi_\Lambda : X \to I^\Lambda$ and $\pi^\Lambda_{\Lambda'} : I^\Lambda \to I^{\Lambda'}$ be the canonical coordinate projections.
- $|\Lambda|$ is the cardinality of Λ.
- For $\nu \in \mathcal{M}$ and $\Lambda \subset \mathbb{Z}$, let $\nu\pi_\Lambda^{-1}$ be the projection of ν to I^Λ, i.e. $\nu\pi_\Lambda^{-1}(U) = \nu(\pi_\Lambda^{-1}U)$ for measurable $U \subseteq I^\Lambda$.
- $\mathcal{I}$ is the family of all intervals $\Lambda = [a, b] \subset \mathbb{Z}$ including the empty set.
- $L^1_{\mathbb{Z}} = \{\nu \in \mathcal{M} : \nu\pi_\Lambda^{-1}$ is absolutely continuous w.r.t. m^Λ for all $\Lambda \in \mathcal{I}\}$.
- For $\nu \in L^1_{\mathbb{Z}}$ and $\Lambda \in \mathcal{I}$ we denote by ν_Λ the density of $\nu\pi_\Lambda^{-1}$ w.r.t. m^Λ. If $\Lambda = \emptyset$, then ν_Λ has the constant value $\nu(X)$.
- $BV_{\mathbb{Z}} = \{\nu \in L^1_{\mathbb{Z}} : \mathrm{Var}_{I^\Lambda}(\nu_\Lambda) < \infty$ for all $\Lambda \in \mathcal{I}\}$.

We define two scales of norms on $L^1_{\mathbb{Z}}$ and (subspaces of) $BV_{\mathbb{Z}}$.

Definition 4.1. *For $0 < \theta \le 1$ and $\nu \in L^1_{\mathbb{Z}}$ let*

$$|\nu|_\theta = \sup_{\Lambda\in\mathcal{I}} \theta^{|\Lambda|} \int |\nu_\Lambda|\, dm \tag{62}$$

$$\|\nu\|_\theta = \sup_{\Lambda\in\mathcal{I}} \theta^{|\Lambda|}\, \mathrm{Var}(\nu_\Lambda)\ . \tag{63}$$

(Here $\int |\nu_\Lambda|\, dm$ and $\mathrm{Var}(\nu_\Lambda)$ are shorthand notations for $\int_{I^\Lambda} |\nu_\Lambda|\, dm^\Lambda$ and $\mathrm{Var}_{I^\Lambda}(\nu_\Lambda)$, respectively.) Observe that $|\nu|_\theta \le \frac{1}{2}\|\nu\|_\theta$ in view of (44).

L^1_θ and BV_θ are now defined to be the completions of $L^1_{\mathbb{Z}}$ and of the space $\{\nu \in BV_{\mathbb{Z}} : \|\nu\|_\theta < \infty\}$, w.r.t. the norms[15] $|.|_\theta$ and $\|.\|_\theta$, respectively.

[14] For a proof of this let $\psi : I \to \mathbb{R}$ be any bounded measurable function. By the law of large numbers $\frac{1}{n}\sum_{k=0}^{n-1} \psi(x_k)$ converges to $\int_I \psi\, dm$ for $m^{\mathbb{Z}}$-a.e. $\boldsymbol{x}$ and to $\int_I \psi\, d\mu$ for $\mu^{\mathbb{Z}}$-a.e. $\boldsymbol{x}$. It follows that $\mu^{\mathbb{Z}}$ is absolutely continuous w.r.t. $m^{\mathbb{Z}}$ if and only if these two integrals coincide for any such function ψ, i.e. if $\mu = m$. We note for later use that the same argument applies to any two stationary product measures on $I^{\mathbb{Z}}$. In particular, two such measures are singular to each other if they are not identical.

[15] $|.|_\theta$ and $\|.\|_\theta$ are obviously seminorms. To see that $|.|_\theta$ is indeed a norm, suppose that $|\nu|_\theta = 0$ for some $\nu \in L^1_{\mathbb{Z}}$. Then $\nu_\Lambda = 0$ for all $\Lambda \in \mathcal{I}$ so that $\nu(\varphi) = 0$ for each $\varphi \in C(X)$ which depends on only finitely many coordinates. As the space of these functions is dense in $C(X)$, this means that $\nu = 0$ as a signed measure. As $|\nu|_\theta \le \|\nu\|_\theta$ by (44), also $\|.\|_\theta$ is a norm.

Example 4.1. Let f be a probability density on I and consider the infinite stationary product measure ν with one-dimensional factors fm. Observe that $\mathrm{Var}_{I^\Lambda}(f) \geq 2 = \mathrm{Var}_{I^\Lambda}(1)$ for all Λ by (44). Hence $|\nu|_\theta = 1$ and $\|\nu\|_\theta = \mathrm{Var}_I(f)$ for all θ. In particular, $\|m^{\mathbb{Z}}\|_\theta = 2$ for all θ.

In the following lemma we collect a few simple observations.

Lemma 4.1. *(a)* $\mathrm{Var}_{I^{\Lambda'}}(\nu_{\Lambda'}) \leq \mathrm{Var}_{I^\Lambda}(\nu_\Lambda)$ *for* $\nu \in L^1_{\mathbb{Z}}$ *and* $\Lambda' \subset \Lambda \in \mathcal{I}$.
(b) $|\nu|_{\theta=1} = \lim_{n\to\infty} \int |\nu_{[-n,n]}|\, dm$ *and* $\|\nu\|_{\theta=1} = \lim_{n\to\infty} \mathrm{Var}(\nu_{[-n,n]})$.
(c) $|\nu|_\theta \leq |\nu|_{\theta=1}$ *and* $\|\nu\|_\theta \leq \|\nu\|_{\theta=1}$ *for all* $\theta \in (0,1]$ *and all* $\nu \in L^1_{\mathbb{Z}}$.

Proof. (a) Just observe that if $\varphi \in \mathcal{T}_{I^{\Lambda'}}$, then $\varphi \circ \pi^\Lambda_{\Lambda'} \in \mathcal{T}_{I^\Lambda}$, and for each $i \in \Lambda'$

$$\int \nu_{\Lambda'}\, \partial_i \varphi\, dm^{\Lambda'} = \int (\nu_{\Lambda'} \circ \pi^\Lambda_{\Lambda'})\, \partial_i(\varphi \circ \pi^\Lambda_{\Lambda'})\, dm^\Lambda = \int \nu_\Lambda\, \partial_i(\varphi \circ \pi^\Lambda_{\Lambda'})\, dm^\Lambda\ .$$

Now (b) follows from (a), and (c) is a direct consequence of the definitions. □

As we are going to describe the quantitative dynamical properties of coupled systems in terms of properties of transfer operators acting on the spaces L^1_θ and BV_θ, it is worth to spend some effort to give more concrete models of these spaces which are defined rather abstractly as completions.

Remark 4.1. $L^1_{\theta=1} = L^1_{\mathbb{Z}}$

It is a little exercise in measure theory to see that, for $\nu \in L^1_{\mathbb{Z}}$, $|\nu|_{\theta=1} = \sup_{\Lambda\in\mathcal{I}} \int |\nu_\Lambda|\, dm$ coincides with the total variation norm $|\nu|_1$ of the signed measure ν.[16] Hence $L^1_{\theta=1}$ is just the closed subspace $L^1_{\mathbb{Z}}$ of $(\mathcal{M}, |.|_1)$.

Remark 4.2. $BV_{\theta=1} = \{\nu \in BV_{\mathbb{Z}} : \sup_{\Lambda\in\mathcal{I}} \mathrm{Var}(\nu_\Lambda) < \infty\}$.

This means that in the definition of $BV_{\theta=1}$ the completion was not necessary. The completeness of the space on the right hand side for the norm $\|.\|_{\theta=1}$ follows easily from the completeness of the spaces $(BV(I^\Lambda), \mathrm{Var}_{I^\Lambda})$.[17]

Hence, for $\theta = 1$, we have defined spaces of signed measures with additional regularity properties, and we will show that our coupled T_ϵ always has a unique invariant measure that belongs to $BV_{\theta=1}$. But neither $\|.\|_{\theta=1}$ nor $|.|_{\theta=1}$ is suited to describe the convergence of measures $P^n_{T_\epsilon}\nu$ to the invariant measure – not even for $\epsilon = 0$ – as the following example shows.

[16] Each signed measure ν has a unique decomposition $\nu = \nu^+ - \nu^-$ as a difference of two finite positive measures which are singular to each other. The total variation norm of ν is defined as $|\nu|_1 = \nu^+(X) + \nu^-(X)$. For a proof that $|\nu|_1 = |\nu|_{\theta=1}$ see e.g. [6, Lemma 2.4].

[17] Let (ν_n) be a Cauchy sequence in $BV_{\theta=1}$. As $|.|_{\theta=1} \leq \|.\|_{\theta=1}$, it is a fortiori a Cauchy sequence in $L^1_{\theta=1}$. Let $\nu = L^1_{\theta=1}\text{-}\lim_{n\to\infty} \nu_n$, and let $\varepsilon_n = \sup_{k\geq n} \|\nu_k - \nu_n\|_\theta$. Let $\Lambda \in \mathcal{I}$, $\varphi \in \mathcal{T}_{I^\Lambda}$, and $i \in \Lambda$. Then $\int (\nu - \nu_n)_\Lambda\, \partial_i\varphi\, dm \leq \lim_k \int (\nu - \nu_k)_\Lambda\, \partial_i\varphi\, dm + \sup_{k\geq n} \int (\nu_k - \nu_n)_\Lambda\, \partial_i\varphi\, dm \leq \varepsilon_n$ so that $\|\nu - \nu_n\|_{\theta=1} = \sup_{\Lambda\in\mathcal{I}} \mathrm{Var}((\nu - \nu_n)_\Lambda) \leq \varepsilon_n \to 0$.

Example 4.2. Let h be the unique invariant probability density of the local p.w.e. map τ, and let μ be the infinite product measure $(hm)^{\mathbb{Z}}$. Recall from Footnote 14 at the beginning of this chapter that any two stationary product measures are singular to each other if they are not identical. Hence, as long as $P_\tau^n 1 \neq h$, the measures $m^{\mathbb{Z}} T_0^{-n}$ and μ are mutually singular, which means that $\|m^{\mathbb{Z}} T_0^{-n} - \mu\|_{\theta=1} \geq |m^{\mathbb{Z}} T_0^{-n} - \mu|_1 = 2$.[18]

We conclude that the "$\theta = 1$"-norms are unsuited to describe the convergence of $m^{\mathbb{Z}} T_0^{-n}$ to the invariant limit measure μ. For this purpose we will use the norms $|.|_\theta$ and $\|.\|_\theta$ with $0 < \theta < 1$. As long as this is all we want to do with them we need not bother that the spaces L^1_θ and BV_θ for $0 < \theta < 1$ are no longer spaces of signed measures. Note, however, that the positive elements in L^1_θ are finite measures on X.

Example 4.3. Here is an example of a Cauchy sequence in L^1_θ whose limit cannot be interpreted as a finite signed measure on X. Let $f : I \to \mathbb{R}$, $\int f\, dm = 0$, $\int |f|\, dm = 1$. Denote by ν_k the infinite product signed measure with one-dimensional factor measures fm at sites $i = 1, \dots, k$ and m at all other sites. Then $|\nu_k|_1 = 1$ and $\nu_k \pi_\Lambda^{-1} = 0$ if $\Lambda \cap \{1, \dots, k\} \neq \emptyset$.

For $r \in (1, \theta^{-1})$ let $\mu_n = \sum_{k=1}^n \alpha_k \nu_k$ with coefficients $|\alpha_k| \leq \frac{r-1}{r} r^k$. Then $\mu_n \pi^{-1}_{\{1,\dots,d\}} = \sum_{k=1}^{d\wedge n} \alpha_k \nu_k$ so that, for each $\Lambda \in \mathcal{I}$ which contains $\{1, \dots, d\}$, $|\mu_n \pi_\Lambda^{-1}|_1 \leq \sum_{k=1}^{d\wedge n} \alpha_k \leq r^d$. Therefore $|\mu_n|_\theta = \sup_{\Lambda\in\mathcal{I}} \theta^{|\Lambda|} |\mu_n \pi_\Lambda^{-1}|_1 \leq 1$. Similarly one shows that, if $l > n$ and $\Lambda \supseteq \{1, \dots, d\}$, then $\theta^{|\Lambda|} |(\mu_n - \mu_l)\pi_\Lambda^{-1}|_1 \leq (\theta r)^n$, whence $|\mu_m - \mu_l|_\theta \leq (\theta r)^n$, and $(\mu_n)_n$ is indeed a Cauchy sequence in L^1_θ. Now let $p \in \mathbb{N}$ such that $r^p > \frac{2r}{r-1}$, and let $\alpha_k = r^k$ if k is an integer multiple of p and $\alpha_k = 0$ otherwise. Then similar estimates show that the total variation norm $|\mu_n|_1$ is at least $\frac{r-1}{r} r^n$, so no "reasonable" limit of the sequence $(\mu_n)_n$ can be a finite signed measure.

4.2 The Infinite Coupled Map

From now on we consider exclusively the case $L = \mathbb{Z}$. Our basic assumptions on τ and Φ_ϵ are the same as those made in Sect. 3.1, namely

- $X = I^{\mathbb{Z}}$ is the state space of the system.
- $\tau : I \to I$ is a p.w.e. map as defined in Sect. 2.1.
- $T_0 : X \to X$ is the infinite product of the map τ: $(T_0 \boldsymbol{x})_i = \tau(x_i)$ $(i \in \mathbb{Z})$.
- $\Phi_\epsilon : X \to X$ $(|\epsilon| < \epsilon_0)$ is a family of coupling maps ϵ-close to the identity in a C^2 sense made precise by the notion of (a_1, a_2)-coupling in (36).
- $T_\epsilon = \Phi_\epsilon \circ T_0 : X \to X$ are the maps describing the coupled systems.

We assume additionally

- The Φ_ϵ have *finite coupling range* $w > 0$, i.e. $\partial_j \Phi_{\epsilon,i} = 0$ whenever $|i-j| > w$. So $A'_{ij} = A''_{ij} = 0$ when $|i - j| > w$ for the matrices A', A'' introduced in (36).

[18] If $\nu \in \mathcal{M}$, then $\nu T_\epsilon^{-1}(U) = \nu(T_\epsilon^{-1} U)$, so $\nu T_\epsilon^{-1} \in \mathcal{M}$.

For a proof of the existence of a T_ϵ-invariant measures in $BV_{\theta=1}$ (Sect. 4.4) one knows how to work around this additional assumption, see [7], but we need a finite coupling range for the proof of a spectral gap in Sect. 4.7. So we make our life easier using this assumption throughout this chapter. The diffusive nearest neighbor coupling (3), for example, has coupling range $w = 1$.

In order to reduce estimates on transformed measures $\nu T_\epsilon^{-\ell}$ on the infinite system to estimates on their finite-dimensional marginal densities $(\nu T_\epsilon^{-\ell})_\Lambda$, we must relate projections onto different Λ to each other: there are no isolated finite subsystems in infinite coupled systems. To be more specific, for $\Lambda = [a, b] \in \mathcal{I}$ and $\ell \in \mathbb{N}$ let $\Lambda(\ell) = [a - \ell w, b + \ell w]$. Denote $\iota_\Lambda : I^\Lambda \to X$ the map $(\iota_\Lambda(\boldsymbol{x}))_i = x_i$ if $i \in \Lambda$ and $(\iota_\Lambda(\boldsymbol{x}))_i = 0$ otherwise[19], and let

$$T_{\epsilon,\Lambda} = \Phi_{\epsilon,\Lambda} \circ T_{0,\Lambda} \quad \text{where } T_{0,\Lambda} = \pi_\Lambda \circ T_0 \circ \iota_\Lambda,\ \Phi_{\epsilon,\Lambda} = \pi_\Lambda \circ \Phi_\epsilon \circ \iota_\Lambda\ . \tag{64}$$

Observe that $T_{0,\Lambda}$ is just the uncoupled map on I^Λ and that $\Phi_{\epsilon,\Lambda}$ is a (a_1, a_2)-coupling on the finite-dimensional space I^Λ with the same constants a_1, a_2 as above. Hence all considerations of Chap. 3, in particular the Lasota–Yorke inequality (55), apply to $T_{\epsilon,\Lambda}$. The important link between T_ϵ and $T_{\epsilon,\Lambda}$ is given by

$$\pi_\Lambda \circ T_\epsilon^\ell = \pi_\Lambda^{\Lambda(\ell)} \circ T_{\epsilon,\Lambda(\ell)}^\ell \circ \pi_{\Lambda(\ell)} \quad \text{for all } \Lambda \in \mathcal{I} \text{ and } \ell \in \mathbb{N}. \tag{65}$$

This follows immediately from the finite coupling range property of Φ_ϵ: no influence of a coordinate x_i with $i \in L \setminus \Lambda(\ell)$ can propagate to Λ within ℓ steps of time.

4.3 The Transfer Operator and a Lasota–Yorke Inequality

We are going to define transfer operators P_{T_ϵ} on L^1_θ in terms of the action of T_ϵ on the densities ν_Λ of the finite-dimensional projections of $\nu \in L^1_{\mathbb{Z}}$. Observe that, for $\nu \in L^1_{\mathbb{Z}}$ and $\varphi : I^\Lambda \to \mathbb{R}$, (65) implies

$$\begin{aligned} \int_X \varphi \circ \pi_\Lambda \, d(\nu T_\epsilon^{-\ell}) &= \int_X \varphi \circ \pi_\Lambda \circ T_\epsilon^\ell \, d\nu \\ &= \int_{I^{\Lambda(\ell)}} (\varphi \circ \pi_\Lambda^{\Lambda(\ell)}) \circ T_{\epsilon,\Lambda(\ell)}^\ell \cdot \nu_{\Lambda(\ell)} \, dm^{\Lambda(\ell)}\ . \end{aligned} \tag{66}$$

This means that

$$(\nu T_\epsilon^{-\ell})_\Lambda = \left((P_{T_{\epsilon,\Lambda(\ell)}^\ell} \nu_{\Lambda(\ell)}) m^{\Lambda(\ell)} \right)_\Lambda =: (P_{T_{\epsilon,\Lambda(\ell)}^\ell} \nu_{\Lambda(\ell)})_\Lambda \tag{67}$$

where we take the last term just as a short hand for the middle one. Hence,

$$\begin{aligned} |\nu T_\epsilon^{-\ell}|_\theta &= \sup_{\Lambda \in \mathcal{I}} \theta^{|\Lambda|} \int |(\nu T_\epsilon^{-\ell})_\Lambda| \, dm \le \sup_{\Lambda \in \mathcal{I}} \theta^{|\Lambda|} \int |P_{T_{\epsilon,\Lambda(\ell)}^\ell} \nu_{\Lambda(\ell)}| \, dm \\ &\le \theta^{-2\ell w} \sup_{\Lambda \in \mathcal{I}} \theta^{|\Lambda(\ell)|} \int |\nu_{\Lambda(\ell)}| \, dm \le \theta^{-2\ell w} \, |\nu|_\theta\ . \end{aligned} \tag{68}$$

[19] Any other measurable section from I^Λ to X would do as well.

In particular, $\nu \mapsto \nu T_\epsilon^{-1}$ is a linear operator on $L^1_{\mathbb{Z}}$ bounded w.r.t. the norm $|.|_\theta$. Hence it extends to a bounded linear operator P_{T_ϵ} on L^1_θ. That P_{T_ϵ} is also a bounded linear operator on BV_θ is an immediate consequence of the following Lasota–Yorke type inequality.

Proposition 4.1 (Lasota–Yorke inequality). *Let $\ell \in \mathbb{N}$. For each $\alpha_\ell > \frac{2}{\kappa_\ell}$ and each $C_{4,\ell} > D_\ell + E_\ell$ there are $\epsilon_1 \in (0, \epsilon_0]$ and $\theta_1 \in (0,1)$ such that for $|\epsilon| \le \epsilon_1$, $\theta \in [\theta_1, 1]$, and $\nu \in BV_\theta$,*

$$|P^\ell_{T_\epsilon} \nu|_\theta \le \theta^{-2w\ell} |\nu|_\theta \tag{69}$$

$$\|P^\ell_{T_\epsilon} \nu\|_\theta \le \alpha_\ell \, \|\nu\|_\theta + C_{4,\ell} \, |\nu|_\theta \, . \tag{70}$$

Given ℓ, α_ℓ and $C_{4,\ell}$, the choice of θ_1 depends only on the coupling range w, that of ϵ_1 only on the constants a_1, a_2 which qualify Φ_ϵ as a (a_1, a_2)-coupling, see (36).

Observe the difference between (69) and the corresponding inequality (54) for finite systems, where P_{T_ϵ} is a contraction w.r.t. the weak norm.

Proof. Equation (69) is just a restatement of (68). We turn to (70). Let $\tilde\alpha_\ell = (\frac{2}{\kappa_\ell}\alpha_\ell)^{1/2}$ and $\tilde C_{4,\ell} = ((D_\ell + E_\ell) C_{4,\ell})^{1/2}$. In view of (67), Lemma 4.1a and the finite-dimensional Lasota–Yorke inequality (55), we have for each $\nu \in L^1_{\mathbb{Z}}$

$$\begin{aligned}
\|P^\ell_{T_\epsilon} \nu\|_\theta &= \|\nu T_\epsilon^{-\ell}\|_\theta = \sup_{\Lambda \in \mathcal{I}} \theta^{|\Lambda|} \operatorname{Var}((\nu T_\epsilon^{-\ell})_\Lambda) \\
&= \sup_{\Lambda \in \mathcal{I}} \theta^{|\Lambda|} \operatorname{Var}\left((P^\ell_{T_{\epsilon,\Lambda(\ell)}} \nu_{\Lambda(\ell)})_\Lambda \right) \\
&\le \theta^{-2w\ell} \sup_{\Lambda \in \mathcal{I}} \theta^{|\Lambda(\ell)|} \left(\tilde\alpha_\ell \operatorname{Var}(\nu_{\Lambda(\ell)}) + \tilde C_{4,\ell} \int |\nu_{\Lambda(\ell)}| \, dm \right) \\
&\le \theta^{-2w\ell} \tilde\alpha_\ell \, \|\nu\|_\theta + \theta^{-2w\ell} \tilde C_{4,\ell} \, |\nu|_\theta
\end{aligned}$$

for $|\epsilon| \le \epsilon_1$, where the choice of ϵ_1 depends on $\ell, \alpha_\ell, C_{4,\ell}, a_1$, and a_2, see Proposition 3.2. Now choose θ_1 such that $\theta_1^{-2w\ell} \le \min\{\alpha_\ell/\tilde\alpha_\ell, C_{4,\ell}/\tilde C_{4,\ell}\}$. □

As in Sect. 2.4 it follows that there are constants $C_1', C_2' > 0$ such that

$$\|P^n_{T_\epsilon} \nu\|_\theta \le C_1' \, \alpha^n \, \|\nu\|_\theta + C_2' \theta^{-2wn} \, |\nu|_\theta \le \left(C_1' + \frac{1}{2} C_2' \theta^{-2wn} \right) \|\nu\|_\theta \tag{71}$$

for all $\nu \in BV_\theta$ and all $n \in \mathbb{N}$ provided $|\epsilon| \le \epsilon_1$ and $\theta \in [\theta_1, 1]$. The constant α can be any number in $((\frac{2}{\kappa_M})^{1/M}, 1)$. C_1', C_2', ϵ_1, and θ_1 will then be chosen as indicated above and at the end of Sect. 2.4. (Recall that $\kappa_M > 2$ by assumption (1).)

4.4 Existence of Invariant Measures with Absolutely Continuous Finite-Dimensional Marginals

In this section we prove the existence of (at least) one probability measure in $BV_{\theta=1}$ which is invariant under T_ϵ.

Theorem 4.1. *Let $T_\epsilon : X \to X$ be a coupled map on $X = I^{\mathbb{Z}}$ as described in Sect. 4.2. For each $|\epsilon| \le \epsilon_1$ there exists a T_ϵ-invariant probability measure μ_ϵ which belongs to $BV_{\theta=1}$. Indeed, $\|\mu_\epsilon\|_{\theta=1} \le C_2'$.*

Proof. Let $\nu = m^{\mathbb{Z}}$. As $|\nu|_{\theta=1} = 1$ and $\|\nu\|_{\theta=1} = 2$ (see Example 4.1), we have $\limsup_{n\to\infty} \|P_{T_\epsilon}^\ell \nu\|_{\theta=1} \le C_2'$ by (71). Let $\nu_n = \frac{1}{n}\sum_{k=0}^{n-1} P_{T_\epsilon}^k \nu$. Then also $\limsup_{n\to\infty} \|\nu_n\|_{\theta=1} \le C_2'$. This implies that $\limsup_{n\to\infty} \mathrm{Var}_{I^\Lambda}((\nu_n)_\Lambda) \le C_2'$ for each $\Lambda \in \mathcal{I}$. As BV_{I^Λ} embeds compactly into $L^1_{I^\Lambda}$ (see Sect. 3.4), there is a subsequence of $((\nu_n)_\Lambda)_{n>0}$ which converges in $L^1_{I^\Lambda}$ to some probability density $h_\Lambda \in L^1_{I^\Lambda}$. By a diagonal procedure one even finds such a subsequence for which this convergence holds for all $\Lambda \in \mathcal{I}$. Observe that the family of densities h_Λ, $\Lambda \in \mathcal{I}$, is consistent in the sense that for any $\Lambda' \subset \Lambda$ holds $(h_\Lambda m^\Lambda)(\pi^\Lambda_{\Lambda'})^{-1} = h_{\Lambda'} m^{\Lambda'}$, because all the $(\nu_n)_\Lambda$ have the same property. Hence, by Kolmogorov's theorem, there is a probability measure $\mu_\epsilon \in \mathcal{M}$ such that $\mu_\epsilon \pi_\Lambda^{-1} = h_\Lambda m^\Lambda$ for all $\Lambda \in \mathcal{I}$. As $\lim_{n\to\infty} \int |(\mu_\epsilon)_\Lambda - (\nu_n)_\Lambda|\, dm^\Lambda = 0$, the estimate $\mathrm{Var}_{I^\Lambda}((\mu_\epsilon)_\Lambda) \le \liminf_{n\to\infty} \mathrm{Var}_{I^\Lambda}((\nu_n)_\Lambda) \le C_2'$ follows from (48). Hence $\|\mu_\epsilon\|_{\theta=1} \le C_2'$.

It remains to show that μ_ϵ is T_ϵ-invariant:

$$\begin{aligned} |\mu_\epsilon T_\epsilon^{-1} - \mu_\epsilon|_{\theta=1} &= \lim_{n\to\infty} |\nu_n T_\epsilon^{-1} - \nu_n|_{\theta=1} \\ &\le \lim_{n\to\infty} \frac{1}{n}(|\nu T_\epsilon^{-n}|_{\theta=1} + |\nu|_{\theta=1}) = 0\,. \end{aligned}$$

□

Corollary 4.1 (Finite entropy density). *The T_ϵ-invariant measures μ_ϵ from Theorem 4.1 have an entropy density bounded by $\ln(C_2'/2)$. Indeed, for each $\nu \in BV_{\theta=1}$ and each $\Lambda \in \mathcal{I}$ we have $\int \nu_\Lambda \ln \nu_\Lambda\, dm \le |\Lambda|\, \ln\left(\frac{1}{2}\|\nu\|_{\theta=1}\right)$.*

Proof. Let $\Lambda = [a, b]$, $\Lambda' = [a, b-1]$. Then

$$\begin{aligned} \int_{I^\Lambda} \nu_\Lambda \ln \nu_\Lambda\, dm &= \int_{I^{\Lambda'}} \nu_{\Lambda'} \ln \nu_{\Lambda'} dm + \int_{I^\Lambda} \nu_\Lambda \ln \frac{\nu_\Lambda}{\nu_{\Lambda'}}\, dm \\ &\le \int_{I^{\Lambda'}} \nu_{\Lambda'} \ln \nu_{\Lambda'} dm + \ln \int_{I^\Lambda} \nu_\Lambda \frac{\nu_\Lambda}{\nu_{\Lambda'}}\, dm \end{aligned} \tag{72}$$

by Jensen's inequality, and

$$\begin{aligned} \int_{I^\Lambda} \nu_\Lambda \frac{\nu_\Lambda}{\nu_{\Lambda'}}\, dm &\le \int_{I^{\Lambda'}} \sup_{x_b} \left((\nu_\Lambda)_{\boldsymbol{x}_{\neq b}}(x_b)\right) d\boldsymbol{x}_{\neq b} \\ &\le \frac{1}{2}\int_{I^{\Lambda'}} \mathrm{Var}_I\left((\nu_\Lambda)_{\boldsymbol{x}_{\neq b}}\right) d\boldsymbol{x}_{\neq b} \le \frac{1}{2}\,\mathrm{Var}_{I^\Lambda}(\nu_\Lambda) \le \frac{1}{2}\|\nu\|_{\theta=1}\,. \end{aligned} \tag{73}$$

Applying the same estimate to smaller and smaller boxes one arrives at

$$\int_{I^\Lambda} \nu_\Lambda \ln \nu_\Lambda\, dm \le |\Lambda|\, \ln\left(\frac{1}{2}\|\nu\|_{\theta=1}\right)\,. \tag{74}$$

4.5 Uniqueness and a Spectral Gap – the Uncoupled Case

The compact embedding of BV_{I^Λ} into $L^1_{I^\Lambda}$ for each $\Lambda \in \mathcal{I}$ which we used in the last section does by no means imply a compact embedding of BV_θ into L^1_θ, which would be needed in order to prove the quasi-compactness of P_{T_0} on BV_θ. In fact, we have already seen in Example 4.2 that one cannot expect this operator to be quasi-compact on $BV_{\theta=1}$ – even when $\epsilon = 0$, and one can argue (differently) that quasi-compactness fails also for $\theta \in (0,1)$.

So we will give a more direct proof that $\lim_{n\to\infty} \|P^n_{T_\epsilon}\nu - \mu_\epsilon\|_\theta = 0$ (with exponential speed) whenever $\theta \in [\theta_1, 1)$ and ν is a probability measure in BV_θ. Observe that this implies in particular the uniqueness of the invariant μ_ϵ, even in BV_θ. (See the proof of Corollary 4.4 for details of this argument.) We start with the uncoupled case $\epsilon = 0$ in this section. The proof is made up such that it can be extended to the coupled case in the next section.

Theorem 4.2 (Spectral gap). *Let $\nu \in L^1_\theta$, and assume that $\nu_\emptyset = \nu(X) = 0$. Then*

$$\|P^n_{T_0}\nu\|_\theta \le \frac{C_6}{1-\theta}\,\hat\rho^n\,\|\nu\|_\theta \tag{75}$$

where $C_6 = C_1'(C_1' + \frac{1}{2}C_2') + 2C_2'C_3$ and $\hat\rho = \max\{\alpha,\rho\}^{1/2} \in (0,1)$ are constants derived from (35) and (71).

The proof relies on the following lemma, a variant of which was used for the first time in [8].

Lemma 4.2. *Let $\Lambda', \Lambda \in \mathcal{I}$, $\Lambda = [a,b]$, $\Lambda' = [a, b-1]$, and let $S : I^{\mathbb{Z}\setminus\{b\}} \to I^{\mathbb{Z}\setminus\{b\}}$ be measurable. Suppose that there is some $\tilde\Lambda \in \mathcal{I}$ such that the maps $(S(\boldsymbol{x}_{\neq b}))_j$, $j \in \Lambda'$, depend only on coordinates $i \in \tilde\Lambda$. Consider the map $\tau \times S : I^{\mathbb{Z}} \to I^{\mathbb{Z}}$, $\boldsymbol{x} \mapsto (\tau(x_b), S(\boldsymbol{x}_{\neq b}))$. Then*

$$\int_{I^\Lambda} |(P^\ell_{\tau\times S}\nu)_\Lambda|\,dm \le (2+C_2)\,C_3\,\rho^\ell\,\mathrm{Var}^b_{I^{\tilde\Lambda}}(\nu_{\tilde\Lambda}) + \int_{I^{\Lambda'}} |(P^\ell_{\tau\times S}\nu)_{\Lambda'}|\,dm \tag{76}$$

with constants C_2, C_3, and ρ from (34) and Corollary 2.3.

Proof. Let $\Lambda \in \mathcal{I}$. As $\int_{I^\Lambda} |(P^\ell_{\tau\times S}\nu)_\Lambda|\,dm = \sup_\psi \int_X \psi\, d(P^\ell_{\tau\times S}\nu)$ where the supremum extends over all continuous $\psi : X \to \mathbb{R}$ that depend only on coordinates x_i with $i \in \Lambda$ and satisfy $|\psi| \le 1$, we start by estimating the integrals under the supremum. Given such a test function ψ, let

$$\Psi(\boldsymbol{x}) = \int_0^{x_b} \psi(\tau^\ell(\xi), S^\ell(\boldsymbol{x}_{\neq b}))\,d\xi - x_b \int_0^1 h(\xi)\psi(\tau^\ell(\xi), S^\ell(\boldsymbol{x}_{\neq b}))\,d\xi \tag{77}$$

where h is the unique invariant density of the p.w.e. map τ, see Sect. 2.5. Then

$$\partial_b\Psi(\boldsymbol{x}) = \psi\circ(\tau\times S)^\ell(\boldsymbol{x}) - \bar\psi\circ(\tau\times S)^\ell(\boldsymbol{x}) \tag{78}$$

where $\bar\psi(\boldsymbol{x}) = \int_0^1 h(\xi)\psi(\xi, \boldsymbol{x}_{\neq b})\,d\xi$. Observe that $\bar\psi$ depends only on coordinates x_i with $i \in \Lambda'$. In particular $\bar\psi$ does in fact not depend on x_b. Furthermore, Ψ depends only on x_b and on coordinates x_i with $i \in \tilde\Lambda$. Hence

$$\begin{aligned}\int_X \psi\, d(P^\ell_{\tau\times S}\nu) &= \int_X \psi\circ(\tau\times S)^\ell\, d\nu = \int_X \partial_b\Psi\, d\nu + \int_X \bar\psi\circ(\tau\times S)^\ell\, d\nu \\ &= \int_{I^{\tilde\Lambda}} \partial_b\Psi\, \nu_{\tilde\Lambda}\, dm + \int_{I^{\Lambda'}} \bar\psi\,(P^\ell_{\tau\times S}\nu)_{\Lambda'}\, dm \\ &\le \sup_{\boldsymbol{x}} |\Psi(\boldsymbol{x})|\ \mathrm{Var}^b_{I^{\tilde\Lambda}}(\nu_{\tilde\Lambda}) + \int_{I^{\Lambda'}} |(P^\ell_{\tau\times S}\nu)_{\Lambda'}|\, dm \end{aligned} \tag{79}$$

and we must estimate $\sup_{\boldsymbol{x}} |\Psi(\boldsymbol{x})|$:

$$\Psi(\boldsymbol{x}) = \int_0^1 P^\ell_\tau(1_{[0,x_b]} - x_b\, h)(\xi)\,\psi(\xi, S^\ell(\boldsymbol{x}_{\neq b}))\,d\xi \le C_3\rho^\ell \|1_{[0,x_b]} - x_b\, h\|_{BV} \tag{80}$$

by inequality (35). As $\|h\|_{BV} \le C_2$, (76) follows now from (79) and (80). □

Proof of Theorem 4.2. Let $\Lambda = \tilde\Lambda = [a,b] \in \mathcal{I}$, $\Lambda' = [a, b-1]$, and denote by S the uncoupled map on $I^{\mathbb{Z}\setminus\{b\}}$. Then $T_0 = \tau \times S$, and Lemma 4.2 implies

$$\int_{I^\Lambda} |(P^\ell_{T_0}\nu)_\Lambda|\, dm \le (2 + C_2)\, C_3\, \rho^\ell\ \mathrm{Var}_{I^\Lambda}(\nu_\Lambda) + \int_{I^{\Lambda'}} |(P^\ell_{T_0}\nu)_{\Lambda'}|\, dm\,. \tag{81}$$

Multiplying this inequality by θ^Λ and taking suprema over all $\Lambda \in \mathcal{I}$ this yields

$$|P^\ell_{T_0}\nu|_\theta \le (2 + C_2)\, C_3\, \rho^\ell \|\nu\|_\theta + \theta\, |P^\ell_{T_0}\nu|_\theta \tag{82}$$

where one has to keep in mind that $(P^\ell_{T_0}\nu)_\emptyset = \nu(T_0^{-\ell}X) = \nu(X) = 0$. Hence

$$|P^\ell_{T_0}\nu|_\theta \le \frac{(2 + C_2)\, C_3}{1-\theta}\, \rho^\ell\, \|\nu\|_\theta\,. \tag{83}$$

We combine this with the Lasota–Yorke type estimate (71) for the special case $w = 0$: for all $k, \ell \in \mathbb{N}$,

$$\begin{aligned}\|P^{k+\ell}_{T_0}\nu\|_\theta &\le C_1'\alpha^k \|P^\ell_{T_0}\nu\|_\theta + C_2' |P^\ell_{T_0}\nu|_\theta \\ &\le C_1'\alpha^k (C_1' + \frac{1}{2}C_2')\|\nu\|_\theta + C_2' \frac{(2 + C_2)\, C_3}{1-\theta}\, \rho^\ell\, \|\nu\|_\theta\,. \end{aligned} \tag{84}$$

With $k, l = [\frac{n}{2}](+1)$ and $\hat\rho = \max\{\rho^{1/2}, \alpha^{1/2}\}$ this yields (75). □

Proof of Equation (61). For $j = 1, \dots, d$ let $X_j = I^{\{1,\dots,j\}}$. Let $f \in BV(X_d)$ and define $f_j : X_j \to \mathbb{R}$, $f_j(\boldsymbol{x}_{1:j}) = \int f(\boldsymbol{x})\, dx_{j+1} \dots dx_d$. Then $\mathrm{Var}_{X_j}(f_j) \le \mathrm{Var}_{X_d}(f_d)$ for all $j = 1, \dots, d$, and if $\int f\, dm = 0$, a repeated application of (82) to any measure ν with $\nu_{\{1,\dots,d\}} = f$ yields

$$\int_{X_d} |P^\ell_{T_0} f|\, dm \le (2+C_2)C_3\,\rho^\ell \sum_{j=1}^{d} \mathrm{Var}_{X_j}(f_j) \le (2+C_2)C_3\,\ell\,\rho^\ell\ \mathrm{Var}_{X_d}(f)\ . \quad (85)$$

□

The previous inequality implies in particular that finite uncoupled systems have a unique absolutely continuous invariant probability measure. Indeed, the existence of a P_{T_0}-invariant probability density $h_0 \in BV(X)$ is proved in Theorem 3.1. (Observe that $h_0(\boldsymbol{x}) = h(x_1)\dots h(x_d)$ where h is the unique invariant density for the single site map τ.) Suppose there is another P_{T_0}-invariant probability density $h_0' \in L^1_X$. As $C^1(X)$ is dense in L^1_X there is, for each $\delta > 0$, some $f_\delta \in C^1(X) \subset BV(X)$ with $\int_X |h_0' - f_\delta|\, dm < \delta$. Then (85) implies, for each $\ell > 0$,

$$\begin{aligned}\int |h_0 - h_0'|\, dm &\le \int |P^\ell_{T_0}(h_0 - f_\delta)|\, dm + \int |P^\ell_{T_0}(f_0 - h_0')|\, dm \\ &\le (2+C_2)C_3\,\ell\,\rho^\ell\ \mathrm{Var}_{X_d}(f) + \int |f_0 - h_0'|\, dm\ .\end{aligned}$$

In the limit $\ell \to \infty$ this yields $\int |h_0 - h_0'|\, dm \le \delta$. This proves the claim.

Corollary 4.2. *The infinite uncoupled system has a unique invariant probability measure with absolutely continuous finite-dimensional marginals. (This would not be true for coupled systems as shown by the examples in [26]. See also the chapter by E. Jarvenpää in this book.)*

Proof. If ν is a T_0-invariant probability measure with absolutely continuous finite-dimensional marginal densities ν_Λ, then these densities are invariant for the uncoupled system on I^Λ. By the previous observation, $\nu\pi_\Lambda^{-1}$ is therefore the product measure μ^Λ, where μ is the unique absolutely continuous τ-invariant probability measure. Hence $\nu = \mu^{\mathbb{Z}}$ by Kolmogorov's theorem. □

4.6 A Perturbation Result and a Decoupling Estimate

For our treatment of infinite coupled systems we need a procedure to "decouple" a given site b from all other sites. Technically this boils down to compare a coupling Φ_ϵ with a modified one. Following [9] we provide such an estimate in this section.

Proposition 4.2. *Let $F, \tilde F : X \to X$ be two Lipschitz maps*[20] *with Lipschitz constant $L > 0$ that are close in the following sense: There are constants $K_0, K_1, K_2 > 0$ such that*

[20] $F : X \to X$ is a "Lipschitz map", if all $F_i(\boldsymbol{x})$ are Lipschitz w.r.t. each coordinate x_j with uniformly bounded Lipschitz constants. This means in particular that all partial derivatives of all F_i exist Lebesgue-a.e., are uniformly bounded and that $F_i(x + se_k) - F_i(x) = \int_0^s \partial_k F_i(x + \xi e_k) d\xi$.

(i) $\sum_{i\in\mathbb{Z}} \sup_{\boldsymbol{x}} |\tilde{F}_i(\boldsymbol{x}) - F_i(\boldsymbol{x})| \le K_0$,
(ii) $\sum_{i\in\mathbb{Z}} \sup_{j\in\mathbb{Z}} \sup_{\boldsymbol{x}} |\partial_j \tilde{F}_i(\boldsymbol{x}) - \partial_j F_i(\boldsymbol{x})| \le K_1$, *and*
(iii) $\sup\{\mathrm{Var}(P_{F_{t,\Lambda}} f) : 0 \le t \le 1, \Lambda \in \mathcal{I}, f \in BV(I^\Lambda), \mathrm{Var}_{I^\Lambda}(f) \le 1\} \le K_2$
where $F_{t,\Lambda} = \pi_\Lambda \circ \left(t\tilde{F} + (1-t)F\right) \circ \iota_\Lambda$, *compare (64).*

Assume also that $\partial_j F_i = 0$ *and* $\partial_j \tilde{F}_i = 0$ *if* $|i-j| > w$. *Then*

$$\int_X |(P_{\tilde{F}}\nu)_\Lambda - (P_F\nu)_\Lambda|\, dm \le K_2 \left(K_0 + \frac{1}{2}K_1\right) \mathrm{Var}_{I^{\Lambda(1)}}(\nu_{\Lambda(1)}) \tag{86}$$

for $\Lambda \in \mathcal{I}$ *and* $\nu \in L^1_{\mathbb{Z}}$.

Proof. The maps $F_{t,\Lambda(1)} : I^{\Lambda(1)} \to I^{\Lambda(1)}$ are Lipschitz with Lipschitz constant L, and for any $\psi \in C^1(I^\Lambda)$ with $|\psi| \le 1$,

$$\begin{aligned}
&\int_X \psi\circ\pi_\Lambda\, d(P_{\tilde{F}}\nu) - \int_X \psi\circ\pi_\Lambda\, d(P_F\nu) = \int_X (\psi\circ\pi_\Lambda\circ\tilde{F} - \psi\circ\pi_\Lambda\circ F)\, d\nu \\
&= \int_{I^{\Lambda(1)}} (\psi\circ\pi_\Lambda\circ\tilde{F}\circ\iota_{\Lambda(1)} - \psi\circ\pi_\Lambda\circ F\circ\iota_{\Lambda(1)})\, \nu_{\Lambda(1)}\, dm \\
&= \int_{I^{\Lambda(1)}} \int_0^1 \frac{\partial}{\partial t}(\psi\circ\pi_\Lambda^{\Lambda(1)}(F_{t,\Lambda(1)}(\boldsymbol{x})))\, dt\, \nu_{\Lambda(1)}(\boldsymbol{x})\, d\boldsymbol{x} \\
&= \int_0^1 \sum_{i\in\Lambda} \left(\int_{I^{\Lambda(1)}} \partial_i\psi(\pi_\Lambda^{\Lambda(1)}(F_{t,\Lambda(1)}(\boldsymbol{x}))) \frac{\partial}{\partial t} F_{t,\Lambda(1),i}(\boldsymbol{x})\, \nu_{\Lambda(1)}(\boldsymbol{x})\, d\boldsymbol{x} \right) dt \\
&= \int_0^1 \sum_{i\in\Lambda} \left(\int_{I^{\Lambda(1)}} \partial_i\psi_\Lambda(\boldsymbol{x}) \left(P_{F_{t,\Lambda(1)}} \left(\frac{\partial}{\partial t} F_{t,\Lambda(1),i}\, \nu_{\Lambda(1)} \right) \right)(\boldsymbol{x})\, d\boldsymbol{x} \right) dt
\end{aligned} \tag{87}$$

where $\psi_\Lambda := \psi\circ\pi_\Lambda^{\Lambda(1)}$. As $\psi \in C^1(I^\Lambda)$ is an arbitrary test function with $\sup|\psi| \le 1$, this implies

$$\begin{aligned}
\int_{I^\Lambda} |(P_{\tilde{F}}\nu)_\Lambda - (P_F\nu)_\Lambda|\, dm &\le \int_0^1 \sum_{i\in\Lambda} \mathrm{Var}_{I^{\Lambda(1)}} \left(P_{F_{t,\Lambda(1)}} \left(\frac{\partial}{\partial t} F_{t,\Lambda(1),i}\, \nu_{\Lambda(1)} \right) \right) dt \\
&\le K_2 \sum_{i\in\Lambda} \mathrm{Var}_{I^{\Lambda(1)}} \left((F_{1,\Lambda(1)} - F_{0,\Lambda(1)})_i\, \nu_{\Lambda(1)} \right) .
\end{aligned} \tag{88}$$

Next recall that the variation of multivariate functions is defined as the maximum of the variations over individual coordinates, see (43). Hence Lemma 2.2b, which provides an estimate for the variation of a product of functions of one variable, carries over to the present setting, and we conclude (observing also (44))

$$\int_{I^\Lambda} |(P_{\tilde{F}}\nu)_\Lambda - (P_F\nu)_\Lambda|\, dm \le K_2 \left(K_0 + \frac{1}{2}K_1\right) \mathrm{Var}_{I^{\Lambda(1)}}(\nu_{\Lambda(1)}) . \tag{89}$$

□

The estimate from this proposition does not allow to compare directly Φ_ϵ and $\Phi_0 = \mathrm{Id}$: for any finite lattice L the right hand side of (86) would grow like $|L|$ – because the constants K_0 and K_1 are supposed to bound sums over all lattice sites. On the infinite lattice the right hand side of (86) is thus infinite. Therefore we apply Proposition 4.2 "site by site" and evaluate (86) by aid of the next lemma.

Let $\Lambda \in \mathcal{I}$ and $b \in \Lambda$. In order to decouple site b from all other sites we introduce the following notation: let $\bar\iota_b : I^{\mathbb{Z}} \to I^{\mathbb{Z}}$ be the map $(\bar\iota_b(\boldsymbol{x}))_i = x_i$ if $i \neq b$ and $(\bar\iota_b(\boldsymbol{x}))_b = 0$. Then define $\Phi_{\epsilon,b}, T_{\epsilon,b} : I^{\mathbb{Z}} \to I^{\mathbb{Z}}$,

$$(\Phi_{\epsilon,b}(\boldsymbol{x}))_i = \begin{cases} x_b & \text{if } i = b \\ (\Phi_\epsilon(\bar\iota_b(\boldsymbol{x})))_i & \text{if } i \neq b \end{cases} \qquad \text{and} \qquad T_{\epsilon,b} = \Phi_{\epsilon,b} \circ T_0 \,. \tag{90}$$

Our task is to show that the passage from Φ_ϵ to $\Phi_{\epsilon,b}$ leads to an error (in the sense of Proposition 4.2) of order $|\epsilon|$ independent of the size of Λ (depending heavily on ℓ, though!). Denote by E_b the $\mathbb{Z} \times \mathbb{Z}$ matrix with $(E_b)_{ij} = 1$ if $i = j \neq b$ and $(E_b)_{ij} = 0$ otherwise.

Lemma 4.3. *Let Φ_ϵ be a (a_1, a_2)-coupling.*

(a) $\sum_{i\in\mathbb{Z}} \sup_{\boldsymbol{x}} |(\Phi_{\epsilon,b}(\boldsymbol{x}))_i - (\Phi_\epsilon(\boldsymbol{x}))_i| \le 2|\epsilon|(a_1+1)$.
(b) $\sum_{i\in\mathbb{Z}} \sup_{j\in\mathbb{Z}} \sup_{\boldsymbol{x}} |\partial_j(\Phi_{\epsilon,b}(\boldsymbol{x}))_i - \partial_j(\Phi_\epsilon(\boldsymbol{x}))_i| \le 2|\epsilon|(2a_1+a_2)$.
(c) All $F_t := t\Phi_{\epsilon,b} + (1-t)\Phi_\epsilon$ $(0 \le t \le 1)$ are (a_1, a_2)-couplings and satisfy assumption (iii) in Proposition 4.2 for $K_2 = \frac{1-2a_1|\epsilon|+a_2|\epsilon|}{(1-2a_1|\epsilon|)^2}$, and $K_2 \le 2$ as long as $|\epsilon| \le \min\{\frac{1}{6a_1}, \frac{2}{9a_2}\}$.

Proof. The following estimate yields (a):

$$\begin{aligned} &\sum_{i\in\mathbb{Z}} \sup_{\boldsymbol{x}} |(\Phi_{\epsilon,b}(\boldsymbol{x}))_i - (\Phi_\epsilon(\boldsymbol{x}))_i| \\ &\quad \le \sup_{\boldsymbol{x}} |x_b - (\Phi_\epsilon(\boldsymbol{x}))_b| + \sum_{i\neq b} \sum_{j\in\mathbb{Z}} \sup_{\boldsymbol{x}} |(D\Phi_\epsilon(\boldsymbol{x}))_{ij}| \sup_{\boldsymbol{x}} |(\bar\iota_b(\boldsymbol{x}))_j - x_j| \\ &\quad \le 2|\epsilon| \left(1 + \sum_{i\neq b} A'_{ib}\right) \le 2|\epsilon|\,(1+a_1)\,. \end{aligned}$$

For (b) observe first that

$$(\partial_j(\Phi_{\epsilon,b} - \Phi_\epsilon))_i = \begin{cases} -\partial_j A_{\epsilon,i} & \text{if } i = b \text{ or } j = b \\ (\partial_j\Phi_{\epsilon,i}) \circ \bar\iota_b - \partial_j\Phi_{\epsilon,i} & \text{if } i \neq b \text{ and } j \neq b\,. \end{cases}$$

The difference $(\partial_j\Phi_{\epsilon,i}) \circ \bar\iota_b - \partial_j\Phi_{\epsilon,i}$ can be bounded by $\sup_{\boldsymbol{x}} |\partial_b(\partial_j\Phi_{\epsilon,i})(\boldsymbol{x})| = \sup_{\boldsymbol{x}} |\partial_j(\partial_b\Phi_{\epsilon,i})(\boldsymbol{x})| \le 2|\epsilon|A''_{ib}$. Hence

$$\begin{aligned} \sum_{i\in\mathbb{Z}} \sup_{j\in\mathbb{Z}} |(\partial_j(\Phi_{\epsilon,b} - \Phi_\epsilon))_i| &\le 2|\epsilon| \left(\sum_{i\in\mathbb{Z}} \sup_{j\in\mathbb{Z}} (A' - E_bA'E_b)_{ij} + \sum_{i\in\mathbb{Z}} (E_bA''E_b)_{ib} \right) \\ &\le 2|\epsilon|(2a_1+a_2)\,. \end{aligned}$$

We turn to (c): That all F_t (and hence also all $F_{t,\Lambda}$) are (a_1, a_2)-couplings is a trivial observation. The K_2-bound then follows from Lemma 3.3. □

Combining Proposition 4.2 and Lemma 4.3 we obtain

Corollary 4.3. *Let Φ_ϵ be a (a_1, a_2)-coupling, $\Lambda \in \mathcal{I}$, and $b \in \Lambda$. Then*

$$\int_X |((P_{\Phi_{\epsilon,b}} - P_{\Phi_\epsilon})\nu)_\Lambda| \, dm \le |\epsilon| \, (8a_1 + 2a_2 + 4) \, \mathrm{Var}_{I^{\Lambda(1)}}(\nu_{\Lambda(1)}) \tag{91}$$

for $\nu \in L^1_{\mathbb{Z}}$, as long as $|\epsilon| \le \min\{\frac{1}{6a_1}, \frac{2}{9a_2}\}$.

4.7 Uniqueness and a Spectral Gap – the Coupled Case

We are going to follow essentially the same strategy as in the proof of Theorem 4.2. To this end we decouple the dynamics at a site $b \in \Lambda = [a, b]$ from all other sites. As a result we obtain an estimate like (81) with an additional error term. Finally we use the Lasota–Yorke inequality to control this error term.

Recall that $\epsilon_1 \in (0, \epsilon_0]$ and $\theta_1 \in (0, 1)$ were determined in Proposition 4.1 and $\hat\rho = \max\{\rho, \alpha\}^{1/2} < 1$ in Theorem 4.2.

Theorem 4.3 (Spectral gap). *Let $\gamma \in (\hat\rho, 1)$. There is $\theta_2 \in [\theta_1, 1)$ such that for each $\theta \in [\theta_2, 1)$ there exist $C_\theta > 0$ and $\epsilon_\theta \in (0, \epsilon_1]$ such that*

$$\|P^n_{T_\epsilon}\nu\|_\theta \le C_\theta \, \gamma^n \, \|\nu\|_\theta \tag{92}$$

for all $|\epsilon| \le \epsilon_\theta$, all $\nu \in L^1_\theta$ with $\nu_\emptyset = \nu(X) = 0$, and all $n \in \mathbb{N}$.

In particular, if $m^{\mathbb{Z}}$ denotes the product Lebesgue measure on X, then

$$\|P^n_{T_\epsilon} m^{\mathbb{Z}} - \mu_\epsilon\|_\theta \le C'_\theta \, \gamma^n := (2 + C'_2) C_\theta \, \gamma^n \, . \tag{93}$$

Before we prove this theorem, we note the following corollary.

Corollary 4.4 (Uniqueness). *Let $\theta \in [\theta_2, 1)$ and $|\epsilon| \le \epsilon_\theta$. There is a unique T_ϵ-invariant probability measure μ_ϵ in BV_θ, and μ_ϵ belongs in fact to $BV_{\theta=1}$.*

Proof. The existence of $\mu_\epsilon \in BV_{\theta=1}$ was proved in Theorem 4.1. For the uniqueness assume that $\tilde\mu_\epsilon \in BV_\theta$ is also T_ϵ-invariant. Let $\nu = \mu_\epsilon - \tilde\mu_\epsilon$. Then Theorem 4.3 applies to ν, and as $P_{T_\epsilon}\nu = \nu$, it follows that $\nu = 0$.

Proof of Theorem 4.3. Let $\Lambda = [a, b]$. The proof consists of three steps. In view of Corollary 4.3 we may first replace $P^\ell_{T_\epsilon}\nu$ by $P^\ell_{T_{\epsilon,b}}\nu$ at the expense of an error of size $\mathcal{O}(\ell|\epsilon|)$. Then we can profit from the product structure of $T_{\epsilon,b}$ (the site b is completely decoupled now!) and reduce estimates on the box Λ to estimates on the smaller box $\Lambda' = [a, b-1]$ as we did it in the proof of Theorem 4.2. Finally we must show that it is indeed sufficient to do all this for a large but fixed ℓ.

To begin with,

$$\begin{aligned}
&\left|\int_{I^\Lambda} |(P^\ell_{T_{\epsilon,b}}\nu)_\Lambda|\,dm - \int_{I^\Lambda} |(P^\ell_{T_\epsilon}\nu)_\Lambda|\,dm\right| \le \int_{I^\Lambda} |(P^\ell_{T_{\epsilon,b}}\nu - P^\ell_{T_\epsilon}\nu)_\Lambda|\,dm\\
&\le \sum_{i=0}^{\ell-1}\int_{I^\Lambda}\left|\left(P^i_{T_{\epsilon,b}}(P_{T_{\epsilon,b}}-P_{T_\epsilon})P^{\ell-i-1}_{T_\epsilon}\nu\right)_\Lambda\right| dm\\
&\le \sum_{i=0}^{\ell-1}\int_{I^{\Lambda(i)}} |((P_{\Phi_{\epsilon,b}}-P_{\Phi_\epsilon})P_{T_0}P^{\ell-i-1}_{T_\epsilon}\nu)_{\Lambda(i)}|\,dm\\
&\le |\epsilon|(8a_1+2a_2+4)\sum_{i=0}^{\ell-1}\mathrm{Var}_{I^{\Lambda(i+1)}}((P_{T_0}P^{\ell-i-1}_{T_\epsilon}\nu)_{\Lambda(i+1)})\\
&\le |\epsilon|(8a_1+2a_2+4)\sum_{i=0}^{\ell-1}\theta^{-|\Lambda(i+1)|}\|P_{T_0}P^{\ell-i-1}_{T_\epsilon}\nu\|_\theta\\
&\le \ell\,\theta^{-|\Lambda(\ell)|}|\epsilon|(8a_1+2a_2+4)\cdot(C_1'+\frac{1}{2}C_2)\|\nu\|_\theta =: \ell\,\theta^{-|\Lambda(\ell)|}|\epsilon|\,C_7\,\|\nu\|_\theta
\end{aligned} \tag{94}$$

where we used Corollary 4.3 and (71).

Exactly the same reasoning yields

$$\left|\int_{I^{\Lambda'}} |(P^\ell_{T_{\epsilon,b}}\nu)_{\Lambda'}|\,dm - \int_{I^{\Lambda'}} |(P^\ell_{T_\epsilon}\nu)_{\Lambda'}|\,dm\right| \le \ell\,\theta^{-|\Lambda(\ell)|}|\epsilon|\,C_7\,\|\nu\|_\theta\,. \tag{95}$$

In the next step we will profit from the decoupling: as $T_{\epsilon,b}: X\to X$ is the direct product $\tau\times S$ of τ with a map $S: I^{\mathbb{Z}\setminus\{b\}}\to I^{\mathbb{Z}\setminus\{b\}}$ for which $(S(\boldsymbol{x}_{\neq b}))_j$, $j\in\Lambda'$ depends only on coordinates from $\tilde\Lambda=[a-\ell w, b+\ell w]$, we can apply Lemma 4.2 again and obtain

$$\int_{I^\Lambda} |(P^\ell_{T_{\epsilon,b}}\nu)_\Lambda|\,dm \le (2+C_2)\,C_3\,\rho^\ell\ \mathrm{Var}_{I^{\tilde\Lambda}}(\nu_{\tilde\Lambda}) + \int_{I^{\Lambda'}} |(P^\ell_{T_{\epsilon,b}}\nu)_{\Lambda'}|\,dm\,. \tag{96}$$

Before we put together (94)–(96) we let $\tilde\gamma=(\hat\rho\gamma)^{1/2}$. Then $\hat\rho<\tilde\gamma<\gamma$ and $\hat\rho/\tilde\gamma=\tilde\gamma/\gamma<1$. So we can fix $\theta_2\in[\theta_1,1)\cap((\tilde\gamma/\gamma)^{1/w},1)$, and for $\theta\in[\theta_2,1)$ we can first choose ℓ_θ such that

$$(2+C_2)C_3\rho^{\ell_\theta}\le(1-\theta)\tilde\gamma^{2\ell_\theta}\theta^{2\ell_\theta w} \text{ and } \left(C_1'\left(C_1'+\frac{1}{2}C_2'\right)+C_2'\right)(\tilde\gamma\theta_2^{-w})^{2\ell_\theta}\le\gamma^{2\ell_\theta} \tag{97}$$

and then $\epsilon_\theta\in(0,\epsilon_1]$ so small that $2\ell_\theta\theta^{-2\ell_\theta w}\epsilon_\theta C_7\le(1-\theta)\tilde\gamma^{2\ell_\theta}$. Then

$$\begin{aligned}
&\theta^{|\Lambda|}\int_{I^\Lambda}|(P^{\ell_\theta}_{T_\epsilon}\nu)_\Lambda|\,dm\\
&\quad\le(1-\theta)\tilde\gamma^{2\ell_\theta}\left(\|\nu\|_\theta+\theta^{|\tilde\Lambda|}\,\mathrm{Var}_{I^{\tilde\Lambda}}(\nu_{\tilde\Lambda})\right)+\theta^{|\Lambda|}\int_{I^{\Lambda'}}|(P^{\ell_\theta}_{S_\epsilon}f)_{\Lambda'}|\,dm
\end{aligned} \tag{98}$$

for $|\epsilon|\le\epsilon_\theta$. Taking the supremum over all $\Lambda\in\mathcal{I}$ this yields, as in (82),

$$|P_{T_\epsilon}^{\ell_\theta}\nu|_\theta \le (1-\theta)\tilde{\gamma}^{2\ell_\theta}\|\nu\|_\theta + \theta\,|P_{T_\epsilon}^{\ell_\theta}\nu|_\theta\ . \tag{99}$$

Now the proof of the proposition can be finished along the lines of the proof of Theorem 4.2: invoking the Lasota–Yorke type estimates (69) and (71) one obtains the exponential estimate

$$\begin{aligned}\|P_{T_\epsilon}^{2\ell_\theta}\nu\|_\theta &\le C_1'\alpha^{\ell_\theta}\|P_{T_\epsilon}^{\ell_\theta}\nu\|_\theta + C_2'\theta^{-2w\ell_\theta}|P_{T_\epsilon}^{\ell_\theta}\nu|_\theta \\ &\le \left(C_1'\tilde{\gamma}^{2\ell_\theta}(C_1' + \frac{1}{2}C_2'\theta^{-2w\ell_\theta}) + C_2'\theta^{-2w\ell_\theta}\tilde{\gamma}^{2\ell_\theta}\right)\|\nu\|_\theta \\ &\le \gamma^{2\ell_\theta}\,\|\nu\|_\theta\ .\end{aligned} \tag{100}$$

This yields $\|P_{T_\epsilon}^n\nu\|_\theta \le \gamma^n\|\nu\|_\theta$ for even multiples $n = 2k\ell_\theta$ of ℓ_θ, valid for $|\epsilon| \le \epsilon_\theta$. For general $n = 2k\ell_\theta + j$ with $0 \le j < 2\ell_\theta$ one uses (71) to conclude that $\|P_{T_\epsilon}^n\nu\|_\theta \le (C_1' + C_2'\theta^{-2w\ell_\theta})\|P_{T_\epsilon}^{2k\ell_\theta}\nu\|_\theta \le \left((C_1' + C_2'\theta^{-2w\ell_\theta})\gamma^{-2\ell_\theta}\right)\gamma^n\|\nu\|_\theta$. This is (92). □

4.8 Exponential Decay of Correlations

Lemma 4.4 (Exponential decay in time). *Let $\phi, \psi : X \to \mathbb{R}$ be bounded observables that depend only on coordinates in intervals Λ_ϕ and Λ_ψ, respectively. Let ϵ and θ be as in Theorem 4.3, and let $\tilde{\mu}$ be a probability measure from $BV_{\theta=1}$. Then*

$$\left|\int \phi\cdot\psi\circ T_\epsilon^n\,d\tilde{\mu} - \int\phi\,d\tilde{\mu}\cdot\int\psi\,d\mu_\epsilon\right| \le C_\theta\theta^{-|\Lambda_\psi|}\,\gamma^n\|\tilde{\mu}\|_{\theta=1}\,\|\phi\|_{C^1}\|\psi\|_{C^0}\ . \tag{101}$$

For $\tilde{\mu} = \mu_\epsilon$ this is slightly stronger than (4) of Theorem 1.1.

Proof. It suffices to restrict to the case where $\int\phi\,d\tilde{\mu} = 0$. Let $\nu = \phi\tilde{\mu}$. Then $\nu(X) = \int\phi\,d\tilde{\mu} = 0$ so that

$$\begin{aligned}\left|\int\phi\cdot\psi\circ T_\epsilon^n\,d\tilde{\mu}\right| &\le \int_{I^{\Lambda_\psi}}|(P_{T_\epsilon}^n\nu)_{\Lambda_\psi}|\,\psi\,dm \le \|\psi\|_{C^0}\,\theta^{-|\Lambda_\psi|}\|P_{T_\epsilon}^n\nu\|_\theta \\ &\le \|\psi\|_{C^0}\,\theta^{-|\Lambda_\psi|}C_\theta\gamma^n\|\nu\|_\theta \le \|\psi\|_{C^0}\,\theta^{-|\Lambda_\psi|}C_\theta\gamma^n\|\nu\|_{\theta=1}\end{aligned} \tag{102}$$

by Theorem 4.3 and Lemma 4.1c when $|\epsilon| \le \epsilon_\theta$.

It remains to bound $\|\nu\|_{\theta=1}$: Remembering Lemma 4.1b it suffices to consider boxes $\Lambda \in \mathcal{I}$ which include Λ_ϕ. Let $\varphi \in \mathcal{T}_{I^\Lambda}$. Then

$$\begin{aligned}\int_{I^\Lambda}\partial_i\varphi\,\nu_\Lambda\,dm &= \int_{I^\Lambda}\partial_i(\varphi\phi)\,\tilde{\mu}_\Lambda\,dm - \int_{I^\Lambda}\varphi\,\partial_i\phi\,\tilde{\mu}_\Lambda\,dm \\ &\le \|\phi\|_{C^0}\,\mathrm{Var}(\tilde{\mu}_\Lambda) + \|\partial_i\phi\|_{C^0}\int|\tilde{\mu}_\Lambda|\,dm \le \|\tilde{\mu}\|_{\theta=1}\,\|\phi\|_{C^1}\ .\end{aligned} \tag{103}$$

Hence $\|\nu\|_{\theta=1} \le \|\phi\|_{C^1}\|\tilde{\mu}\|_{\theta=1}$. This finishes the proof of (101). □

Lemma 4.5 (Exponential decay in space). *Let $\phi, \psi : X \to \mathbb{R}$ be bounded observables that depend only on coordinates in the interval Λ. Let γ and θ be as in Theorem 4.3, and let θ be sufficiently close to 1 that $\gamma' := \gamma^{\frac{1}{2w}}\theta^{-1} < 1$ (recall that w is the coupling range). Then, for $|\epsilon| \le \epsilon_\theta$,*

$$\left| \int \phi \cdot (\psi \circ \sigma^n)\, d\mu_\epsilon - \int \phi\, d\mu_\epsilon \int \psi\, d\mu_\epsilon \right| \le C' {\gamma'}^{|n|-|\Lambda|} \|\phi\|_{C^0} \|\psi\|_{C^0} \tag{104}$$

where σ is the left shift on $X = I^{\mathbb{Z}}$.

Proof. We may assume that $\int \phi\, d\mu_\epsilon = 0$. Let $\tilde\psi = \psi \circ \sigma^n$. As ϕ and $\tilde\psi$ depend on variables at distance $|n|-|\Lambda|$ at least, it follows that $\phi \circ T_\epsilon^k$ and $\tilde\psi \circ T_\epsilon^k$ depend on disjoint sets of variables for $k = [\frac{|n|-|\Lambda|}{2w}]$. Accordingly, by Theorem 4.3,

$$\begin{aligned}
\left| \int \phi \cdot (\psi \circ \sigma^n)\, d\mu_\epsilon \right| &\le \left| \int \phi\, \tilde\psi\, d(P_{T_\epsilon}^k m^{\mathbb{Z}}) \right| + C'_\theta \gamma^k\, \theta^{-|\Lambda|-|n|}\, \|\phi\|_{C^0} \|\psi\|_{C^0} \\
&= \left| \int (\phi \circ T_\epsilon^k)(\tilde\psi \circ T_\epsilon^k)\, dm^{\mathbb{Z}} \right| + C'_\theta \gamma^k\, \theta^{-|\Lambda|-|n|}\, \|\phi\|_{C^0} \|\psi\|_{C^0} \\
&= \left| \int (\phi \circ T_\epsilon^k)\, dm^{\mathbb{Z}} \right| \cdot \left| \int (\tilde\psi \circ T_\epsilon^k)\, dm^{\mathbb{Z}} \right| + C'_\theta \gamma^k\, \theta^{-|\Lambda|-|n|}\, \|\phi\|_{C^0} \|\psi\|_{C^0} \\
&= \left| \int \phi\, d(P_{T_\epsilon}^k m^{\mathbb{Z}}) \right| \cdot \left| \int \tilde\psi\, d(P_{T_\epsilon}^k m^{\mathbb{Z}}) \right| + C'_\theta \gamma^k\, \theta^{-|\Lambda|-|n|}\, \|\phi\|_{C^0} \|\psi\|_{C^0} \\
&\le 2C'_\theta \gamma^k\, \theta^{-|\Lambda|-|n|}\, \|\phi\|_{C^0} \|\psi\|_{C^0}\,.
\end{aligned} \tag{105}$$

As $\gamma' = \gamma^{\frac{1}{2w}}\theta^{-1}$, (104) follows immediately. □

4.9 μ_ϵ as Unique Physical Measure

The invariant measure μ_ϵ has some properties which qualify it as the unique physical (or observable) measure: it governs a strong law of large numbers (Corollary 1.1), and it is stable under small independent random perturbations. Below we prove the first assertion, and we formulate precisely what we mean by the second one (without providing a proof, though).

Proof of Corollary 1.1 (Strong law of large numbers): Let $f : I \to \mathbb{R}$ be a probability density of bounded variation and let $\nu = (fm)^{\mathbb{Z}}$ the infinite product measure of the probability measure with density f. Then $\nu \in BV_{\theta=1}$, and indeed $\|\nu\|_{\theta=1} = \mathrm{Var}_I(f) < \infty$ by Example 4.1. Let ψ be a C^1 observable that depends only on finitely many coordinates. (It clearly suffices to prove the corollary for such observables, because each continuous observable can be uniformly approximated by them.) In view of [10, Theorem 5.1] it suffices to prove the two following properties:[21]

[21] Given (106) and (107), the proof of the law of large numbers is easy. We more or less copy it from [10]. Let $S_n(\boldsymbol{x}) = \sum_{k=0}^{n-1} \left(\psi(T_\epsilon^k \boldsymbol{x}) - \int \psi \circ T_\epsilon^k\, d\nu\right)$. Then

$$\lim_{k\to\infty} \int \psi \circ T_\epsilon^k \, d\nu = \int \psi \, d\mu_\epsilon \tag{106}$$

$$\sup_{k>0} \sum_{j=0}^{\infty} \left| \int (\psi \circ T_\epsilon^j)(\psi \circ T_\epsilon^k) \, d\nu - \int \psi \circ T_\epsilon^j \, d\nu \int \psi \circ T_\epsilon^k \, d\nu \right| \le C_{\theta,\psi} < \infty \, . \tag{107}$$

Equation (106) is contained in the special case $\tilde{\mu} = \nu$ and $\phi = 1$ of (101). Equation (107) follows also from (101): first replace ψ by $\psi - \int \psi \, d\mu_\epsilon$, i.e. assume that $\int \psi \, d\mu_\epsilon = 0$. Then apply (101) with $\tilde{\mu} = P_{T_\epsilon}^\ell \nu$ and $\ell = |j-k|$ to show that

$$\sum_{j=0}^{\infty} \left| \int (\psi \circ T_\epsilon^j)(\psi \circ T_\epsilon^k) \, d\nu \right| \le C'_{\theta,\psi} < \infty$$

and finally apply (101) with $\tilde{\mu} = \nu$ and $\phi = 1$ to show that

$$\sum_{j=0}^{\infty} \left| \int \psi \circ T_\epsilon^j \, d\nu \int \psi \circ T_\epsilon^k \, d\nu \right| \le \|\psi\|_{C^0} \sum_{j=0}^{\infty} \left| \int \psi \circ T_\epsilon^k \, d\nu \right| \le C''_{\theta,\psi} < \infty \, .$$

For both of these estimates one uses the fact that $\sup_k \|P_{T_\epsilon}^k \nu\|_{\theta=1} < \infty$, see (71).

Remark 4.3. Here is another reason why μ_ϵ should be considered as the unique observable invariant measure: it is stable under independent random perturbations. More precisely, consider a smooth family $(S_\omega)_{\omega\in\mathbb{R}}$ of C^2 maps from $I \to I$, $S_0 = \mathrm{Id}_I$. Let $(\nu_\delta)_{\delta>0}$ be a family of probability measures on $\mathbb{R}$ with C^1 densities supported on $[-\delta, \delta]$, and consider the random process defined by the Markov operator[22]

$\int (\frac{1}{n} S_n)^2 \, d\nu \le \frac{1}{n^2} n C_{\theta,\psi} = \frac{1}{n} C_{\theta,\psi}$ by (107), so that the subsequence $(\frac{1}{k^2} S_{k^2})_{k\ge 1}$ is L^1_ν-summable. In particular, $\frac{1}{k^2} S_{k^2} \to 0$ ν-a.e.. To show the a.e.-convergence of the full sequence, let $m_n = [\sqrt{n}]^2$. The bound $m_n \le n \le m_n + 2\sqrt{n}$ ensures $\frac{m_n}{n} \to 1$ and $|S_n - S_{m_n}| \le 2\sqrt{n} \|\psi\|_{C^0}$. Hence

$$\frac{1}{n} S_n = \frac{m_n}{n} \frac{1}{m_n} S_{m_n} + \frac{S_n - S_{m_n}}{n} \to 0 \quad \text{almost surely as } n \to \infty$$

and $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \psi(T_\epsilon^k \boldsymbol{x}) = \int \psi \, d\mu_\epsilon$ for ν-a.e. $\boldsymbol{x}$ follows from (106).

[22] To convince oneself that this is quite a general way to write down random perturbations consider the case in which the S_ω are defined on the circle $I = \mathbb{R}/\mathbb{Z}$ by $S_\omega(x) = x + \omega$ and where $\nu_\delta(d\omega) = \delta^{-1} q(\delta^{-1}\omega) \, d\omega$. Then, for measures $\mu(dx) = h(x) \, dx$,

$$\begin{aligned} \tilde{P}_{\nu_\delta} \mu(\varphi) &= \int_{\mathbb{R}} \int_I \varphi(x+\omega) h(x) \delta^{-1} q(\delta^{-1}\omega) \, dx \, d\omega \\ &= \int_I \varphi(y) \left(\int_{\mathbb{R}} h(y-\omega) \delta^{-1} q(\delta^{-1}\omega) \, d\omega \right) dy \end{aligned} \tag{108}$$

$$\tilde{P}_{\nu_\delta}\mu(\varphi) := \int_{\mathbb{R}}\int_I \varphi\circ S_\omega \, d\mu \, d\nu_\delta(\omega) \, . \tag{109}$$

We will consider independent random perturbations of this type at each site, that is

$$P_{\nu_\delta}\mu(\varphi) := \int_{\mathbb{R}^{\mathbb{Z}}}\int_I \varphi\circ(\otimes_{i\in\mathbb{Z}}S_{\omega_i})\, d\mu \, \otimes_{i\in\mathbb{Z}} d\nu_\delta(\omega_i) \, . \tag{110}$$

Define now $P_{\epsilon,\delta} := P_{\nu,\delta}P_{T_\epsilon}$. One can show the following (actually, this is a nice exercise for the reader):

- $P_{\epsilon,\delta}$ has a unique invariant probability measure $\mu_{\epsilon,\delta}$, and this measure belongs to $BV_{\theta=1}$.
- $\|\mu_{\epsilon,\delta}-\mu_\epsilon\|_\theta \le C_\theta\,\delta\,(\ln\delta^{-1})^2$, in particular $\lim_{\delta\to 0}\mu_{\epsilon,\delta}=\mu_\epsilon$ in the weak topology.

We finish by proving assertion 3 of Theorem 1.1.

Lemma 4.6. *There exist $\theta_*\in(0,1)$ such that, for each $\theta\in[\theta_*,1)$ and $\epsilon\in(0,\epsilon_\theta]$ (where ϵ_θ is as in Theorem 4.3), there exists $C''_\theta>0$ such that*

$$\|\mu_0-\mu_\epsilon\|_\theta \le C''_\theta\,\epsilon\ln\epsilon^{-1} \, . \tag{111}$$

Proof. By a repeated application of Corollary 4.3, for each $\Lambda\in\mathcal{I}$ and each probability measure $\nu\in BV_{\theta=1}$,

$$\theta^{|\Lambda|}\int_{I^\Lambda}|(P_{T_0}\nu-P_{T_\epsilon}\nu)_\Lambda|\,dm \le D'\,\epsilon\,\theta^{|\Lambda|}|\Lambda|\,|\nu|_{\theta=1} \le D\,\epsilon\,|\nu|_{\theta=1}$$

for some constants $D',D>0$. Thus, observing that $|P_{T_0}|_\theta=1$,

$$|P^n_{T_0}\mu_0-P^n_{T_\epsilon}\mu_0|_\theta \le \sum_{k=0}^{n-1}|(P_{T_0}-P_{T_\epsilon})P^{n-k-1}_{T_\epsilon}\mu_0|_\theta \le nC'''_\theta\epsilon$$

for a suitable constant $C'''_\theta>0$. Hence, in view of (92) and Theorem 4.1,

$$|\mu_\epsilon-\mu_0|_\theta \le |\mu_\epsilon-P^n_{T_\epsilon}\mu_0|_\theta+|P^n_{T_0}\mu_0-P^n_{T_\epsilon}\mu_0|_\theta \le C''''_\theta(\gamma^n+n\epsilon)$$

and (111) follows by choosing n proportional to $\ln\epsilon^{-1}$. □

References

1. P. Glendinning: *Stability, Instability and Chaos: An Introduction to the Theory of Nonlinear Differential Equations* (Cambridge University Press 1994), pp 307–312. 116
2. R. Bowen: Israel J. Math. **28**, 161–168 (1977). 116
3. G. Keller: Monatshefte Math. **94**, 313–333 (1982). 116
4. M.L. Blank: Russian Acad. Sci. Dokl. Math. **47**, 1–5 (1993). 116
5. M. Blank, G. Keller: Nonlinearity **10**, 81–107 (1997). 116

6. G. Keller, M. Künzle: Ergod. Th.& Dynam. Sys. **12**, 297–318 (1992). 117, 130, 131, 135
7. M. Künzle: Invariante Maße für gekoppelte Abbildungsgitter, Dissertation, Universität Erlangen (1993). 117, 130, 131, 137
8. M. Schmitt: BV-spectral theory for coupled map lattices, Dissertation, Universität Erlangen (2003). See also: Nonlinearity **17**, 671–690 (2004). 117, 140
9. G. Keller, C. Liverani: Discrete and Continuous Dynamical Systems **11**, n.2& 3, 325–335 (2004). 117, 142
10. H.H. Rugh: Ann. Sci. Ec. Norm. Sup., 4e série, **35**, 489–535 (2002) 117, 148
11. A. Boyarsky, P. Góra: *Laws of Chaos: Invariant Measures and Dynamical Systems in One Dimension* (Birkhäuser, Boston-Basel-Berlin 1997). 119, 120, 124, 125
12. A. Lasota, M.C. Mackey: *Probabilistic properties of deterministic systems* (Cambridge University Press 1985). 119
13. E. Giusti: *Minimal Surfaces and Functions of Bounded Variation* (Birkhäuser, Boston 1984). 124, 127, 133
14. W.P. Ziemer: *Weakly Differentiable Functions* (Springer Verlag, New York 1989). 124, 127, 128, 133
15. L.E. Evans, R.F. Gariepy: *Measure Theory and fine Properties of Functions* (CRC Press, Boca Raton 1992). 124, 127
16. A. Lasota, J.A. Yorke: Transactions Amer. Math. Soc. **186**, 481–488 (1973). 123
17. C. Ionescu-Tulcea, G. Marinescu: Ann. of Math (2) **52**, 140–147 (1950). 124
18. V. Baladi: *Positive Transfer Operators and Decay of Correlations* (Advanced Series in Nonlinear Dynamics, Vol 16, World Scientific, Singapore 2000). 124
19. H. Hennion: Proceedings of the American Mathematical Society **118**, 627–634 (1993). 124
20. G. Keller, C. Liverani: Ann. Mat. Sc. Norm. Pisa **28**, 141–152 (1999). 125
21. G. Keller: Int. J. Bif. Chaos **9**, 1777–1783 (1999). 125
22. C. Liverani: J. Stat. Physics **78**, 1111–1129 (1995). 125
23. C. Liverani: Nonlinearity **14**, 463–490 (2001). 125
24. G. Keller: Coupled map lattices via transfer operators on functions of bounded variation. In: *Stochastic and Spatial Structures of Dynamical Systems*, ed. by S.J. van Strien, S.M. Verduyn Lunel (North Holland, Amsterdam 1996) pp 71–80. 131
25. G. Keller: Proc. Steklov Inst. Math. **216**, 315–321 (1996). 133
26. E. Järvenpää, M. Järvenpää: Commun. Math. Phys. **220**, 1–12 (2001). 142