

Soft and Biological Matter

Xiang-Yang Liu *Editor*

# Frontiers and Progress of Current Soft Matter Research

 Springer

# Soft and Biological Matter

## Series Editors

David Andelman, School of Physics and Astronomy, Tel Aviv University,  
Tel Aviv, Israel

Wenbing Hu, School of Chemistry and Chemical Engineering, Department of  
Polymer Science and Engineering, Nanjing University, Nanjing, China

Shigeyuki Komura, Department of Chemistry, Graduate School of Science and  
Engineering, Tokyo Metropolitan University, Tokyo, Japan

Roland Netz, Department of Physics, Free University of Berlin, Berlin, Berlin,  
Germany

Roberto Piazza, Department of Chemistry, Materials Science, and Chemical  
Engineering “G. Natta”, Polytechnic University of Milan, Milan, Italy

Peter Schall, Van der Waals-Zeeman Institute, University of Amsterdam,  
Amsterdam, Noord-Holland, The Netherlands

Gerard Wong, Department of Bioengineering, California NanoSystems Institute,  
UCLA, Los Angeles, CA, USA

“Soft and Biological Matter” is a series of authoritative books covering established and emergent areas in the realm of soft matter science, including biological systems spanning all relevant length scales from the molecular to the mesoscale. It aims to serve a broad interdisciplinary community of students and researchers in physics, chemistry, biophysics and materials science.

Pure research monographs in the series, as well as those of more pedagogical nature, will emphasize topics in fundamental physics, synthesis and design, characterization and new prospective applications of soft and biological matter systems. The series will encompass experimental, theoretical and computational approaches. Topics in the scope of this series include but are not limited to: polymers, biopolymers, polyelectrolytes, liquids, glasses, water, solutions, emulsions, foams, gels, ionic liquids, liquid crystals, colloids, granular matter, complex fluids, microfluidics, nanofluidics, membranes and interfaces, active matter, cell mechanics and biophysics.

Both authored and edited volumes will be considered.

More information about this series at <http://www.springer.com/series/10783>

Xiang-Yang Liu  
Editor

# Frontiers and Progress of Current Soft Matter Research

 Springer

*Editor*  
Xiang-Yang Liu  
Research Institution for Biomimetics  
and Soft Matter  
Xiamen University  
Xiamen, China

ISSN 2213-1736 ISSN 2213-1744 (electronic)  
Soft and Biological Matter  
ISBN 978-981-15-9296-6 ISBN 978-981-15-9297-3 (eBook)  
<https://doi.org/10.1007/978-981-15-9297-3>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

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

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

# Preface

The soft matter research started more than one hundred years ago. Over a century, the correlated subjects have been examined extensively. People not only keep updating many new soft matter systems, and renewing the knowledge, but also looking into the new applications arising from new demands across areas of physics, biology and chemistry.

The systems of soft matter share some common characteristics, such as notable thermal fluctuations, multiple metastable states, mesoscopic multi-scale self-assembled structures, entropy-driven order-disorder transitions, macro flexibility. Briefly, these are the systems having “small stimulus, big response” and displaying strong nonlinearities. These characteristics are not so much related to their microstructures (at atomic or molecular levels) but more to their mesoscopic self-assembled structures. These obviously belong to multi-scale complex systems and certainly are good subjects for complex statistic physics research.

Within the different types of soft matter systems, flexible materials have been listed as one of the most important materials in recent years, due to the broad applications to big health and related area. In combination with big data and Ai technologies, flexible materials and the correlated flexible electronics will reshape our living and working styles. In this regard, the research efforts on soft materials are targeting on three main aspects, namely recoverable, multi-functional and biocompatible. The current major attention from the community is then focused on the identification and fabrication of new materials with high performance for use in protonic/electronic devices, environment friendly intelligent building materials with the applications in chemo-catalysis, drug delivery, gene delivery, biological imaging and tissue engineering. remote diagnosis related-fields. In particular, wearable, implantable, bio-degradable/absorbable and injectable flexible devices will exert a huge impact on human health and daily life.

This book is based on the lectures delivered by international experts in the 2019 International Graduate Summer School on Soft Matter and Non-equilibrium Physics. The school covers some fundamental aspects and frontier in non-equilibrium physics and soft matter research. Apart from the basic knowledge on nonlinear statistic physics, dynamics, computer simulations, and main

approaches and emerging systems in soft matter research, the particular attention is also devoted to new conceptual flexible functional materials, i.e. silk meso-molecular materials, molecular gels, liquid crystals, and the enriching areas, i.e. flexible electronics, new types of catalysis, etc. One of the intentions of this book is to start with the structure formation dynamics and the correlation between the structures and macroscopic performance. This lays down the foundation for the mesoscopic materials design and functionalization.

The book evolves from the lecturing style of the school. Therefore, the basic principles and technologies of computer simulations and experimental methods are explained in more detail. Illustrations, tables and videos are included in this textbook to improve the readability. Examples are added to help understanding. It can, therefore, be adopted as a reference book for senior undergraduate students, graduate students, and researchers who are interested in soft matter researches.

I am sincerely indebted to the authors for the great efforts in the composition of the respective chapters that combine timely and comprehensive reviews of current frontiers with the fundamental principles to serve the purpose of this book. My appreciation also extends to Dr. Q. L. Huang from Physics, the Xiamen University for her great effort and kind assistance in soliciting the manuscripts from authors, Dr. M. C. Huang and the whole editorial team of *Springer Nature* for their professionalism throughout the whole editing process. Finally, it is our sincere wish that this book will further stimulate exciting multilateral collaborations among the international scientific communities.



Singapore  
August 2020

Xiang-Yang Liu  
State Distinguished Professor

# Contents

<b>1 Introduction to Nonequilibrium Statistical Physics and Its Foundations</b> .....	1
Lamberto Rondoni	
<b>2 On the Foundational Principles of Statistical Mechanics</b> .....	83
Wei-Mou Zheng	
<b>3 Generalized Onsager Principle and It Applications</b> .....	101
Qi Wang	
<b>4 An Introduction to Emergence Dynamics in Complex Systems</b> .....	133
Zhigang Zheng	
<b>5 Basics of Molecular Modeling and Molecular Simulation</b> .....	197
Chenyu Tang (唐晨宇) and Yanting Wang (王延颢)	
<b>6 Cocoon Silk: From Mesoscopic Materials Design to Engineering Principles and Applications</b> .....	241
Wu Qiu and Xiang-Yang Liu	
<b>7 A Primer on Gels (with an Emphasis on Molecular Gels)</b> .....	299
Richard G. Weiss	
<b>8 Fréedericksz-Like Positional Transition Triggered by An External Electric Field</b> .....	323
Ke Xiao and Chen-Xu Wu	



# Chapter 1

## Introduction to Nonequilibrium Statistical Physics and Its Foundations



Lamberto Rondoni 

**Abstract** Statistical physics is a subject in which determinism and randomness, probabilities and rigorous laws relating material properties of macroscopic objects are finely intertwined. This calls for particular care in the interpretation of results. When this is done, statistical physics, and in particular its nonequilibrium branch, applies very widely to natural phenomena, but also to other fields, in which a statistical treatment is appropriate. In this paper, the fundamental elements and results of this subject, meant to interpret the thermodynamic picture of matter in terms of the atomic hypothesis, are reviewed. These include the Brownian motion; the Fluctuation Dissipation Theorem; the Boltzmann equation; the fluctuation relations; the linear and the non-perturbative response theory. Furthermore, recent applications ranging from active matter to the detection of gravitational waves and to machine learning are briefly summarized. The literature cited in support of the text is supplemented by a list of further readings, meant to fill some of the very many inevitable gaps in the presentation of this subject.

**Keywords** Physical laws · Material properties · Determinism · Stochasticity · Probability · Emergence · Response to perturbations · Fluctuations

### 1.1 Prologue: The Realm of Theories and Measurements

Prepare a cup of Italian coffee, which means quite a small amount of hot water and coffee. Lay it on a wooden table in a dining hall in a city, which is on Earth, which is in the Solar system, in the Milky Way etc. Leave the cup on the table for a while; our senses perceive that its temperature changes. In less than half an hour, we realize the coffee has reached the room temperature. After a few days, the coffee is not there

---

L. Rondoni (✉)

Dipartimento di Matematica, Politecnico di Torino,  
corso Duca degli Abruzzi 24, 10129 Torino, Italy  
e-mail: [lamberto.rondoni@polito.it](mailto:lamberto.rondoni@polito.it)

Sezione di Torino, INFN, Via P. Giuria 1, 10125 Torino, Italy

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021  
X.-Y. Liu (ed.), *Frontiers and Progress of Current Soft Matter Research*,  
Soft and Biological Matter, [https://doi.org/10.1007/978-981-15-9297-3\\_1](https://doi.org/10.1007/978-981-15-9297-3_1)

anymore: although nobody drunk it, it evaporated, leaving only dark stains in the cup. Let some thousand years pass; even the table has deteriorated, and maybe the cup like our city have undergone massive destruction, because of war, earthquakes, hurricanes etc. Several billion years later, the Earth is vaporised by the explosion of the Sun, and still later the Milky Way has undergone a destructive collision with the Andromeda galaxy.

Has our coffee ever reached a stationary state? A state in which its physical properties do not change in time? Has any of the small or large objects that we mentioned ever been in a thermodynamic stationary state? Do stationary states exist? Why is thermodynamic so much concerned with equilibrium states, a special kind of stationary states?

Thermodynamics is a most successful branch of Physics describing macroscopic objects, and Physics is the realm of quantitative—hence mathematical—theories meant to describe and predict the being and becoming of natural phenomena. They are motivated by the irresistible human drive to understand the existent and give life to new realities. The formulation of a correct theory requires and inspires experiments (not just experiences) that involve measurements of relevant quantities, in order to test its validity. Measurements should then provide objective data, that become intelligible within the mathematical framework constituting the theory. Objective means that they do not depend on the observer, i.e. their significance stands for a passer-by whether that person is aware or is not aware of the measurement: if water in a pot is 90 °C hot, those immersing a hand in it will be burned, whether they knew or did not know that a temperature measurement had previously been performed; whether a measurement had been performed or not performed.

The goal of thermodynamics is to investigate the consequences of fundamental principles such as the conservation of energy, the impossibility of perpetual motion, etc. Its second law has been described by Lieb and Yngvason [1] as follows:

The second law of thermodynamics is, without a doubt, one of the most perfect laws in physics. Any reproducible violation of it, however small, would bring the discoverer great riches as well as a trip to Stockholm. The world's energy problems would be solved at one stroke. It is not possible to find any other law (except, perhaps, for super selection rules such as charge conservation) for which a proposed violation would bring more skepticism than this one. Not even Maxwell's laws of electricity or Newton's law of gravitation are so sacrosanct, for each has measurable corrections coming from quantum effects or general relativity. The law has caught the attention of poets and philosophers and has been called the greatest scientific achievement of the nineteenth century.

Describing the behaviour of matter at the scale of our daily life—what we call the macroscopic world—not all kinds of measurements are of thermodynamic concern; only measurements that can be performed with certain tools, following certain protocols are considered. This should be understood, because measurement tools and protocols necessarily affect the picture of reality one gets. Indeed, different measurement tools probe different properties of reality, and yield different descriptions of even the very same objects. Wearing glasses that allow only radiation around 510–530 THz makes the whole world look either yellow or black. Other glasses yield a different picture.

Let us discuss how come objectivity is possible, despite inevitable subjective interference in all possible measurements. Indeed, as measurement tools and protocols are chosen by someone, who may want to investigate a given phenomenon and not others, the objectivity of data may look problematic.

The most familiar measurement tools are afforded by our senses: amazingly complex structures, more accurate and dependable than many artificial sophisticated tools. Nevertheless, our sense are imperfect and may badly lead us into error, like hallucinations or misjudgements. Even when such negative situations do not happen, our senses may still induce different people to draw contradictory conclusions on the very same phenomenon. For instance, *nattō* rotten beans, considered a delicacy in Japan, result disgusting to many westerners.

Measurement tools are extensions of our senses, meant to reduce this degree of subjectivity, by increasing accuracy and resolution. Presently, the duration of a second can be split in  $10^{16}$  parts, the length of a forearm in  $10^{10}$  parts, the weight of a pollen grain in  $10^{18}$  parts. This may look sufficiently accurate to eliminate any ambiguity and source of subjectivity in physical measurements. Nevertheless, given an object of investigation, the observer shall decide which aspect of the object should be subjected to measurement, which tools shall be used, which protocols shall be adopted. Moreover, while it is obvious that different quantities, such as time and mass, require different tools to be measured, it is interesting to observe that a single quantity may require different tools, depending on the scale of observation: one cannot weigh a star with a scale built on Earth, while one such scale can weigh diamonds.

Even agreeing on the quantities to be measured and on the tools to be used, people can still get to totally different conclusions. Let us consider a cloud and use our eyes. A common experience of travellers is that the cloud looks like a white object with a definite shape and extension in the blue sky, but flying closer and closer, the cloud boundaries gradually vanish, and after a while there is only grey: no white, no blue. Which picture is more *real*? The white shape in the blue sky, or the grey that seems to have neither a beginning nor an end? Which of the measurement is more accurate? The one performed before take off, or that performed right inside the cloud?

One could conclude that there are too many points of view on a very same object, hence plenty of room for subjectivity rather than objective science. This would seem to undermine the foundations of Galilean science, which has led progress for centuries, even before the recent attacks from big data and machine learning got to the stage. But this is not the case. As Chinese culture teaches, the truth needs the apparently contrasting views to be harmonized. There is room for subjectivity in deciding what is interesting to measure: we may choose to classify the colors of objects, rather than their masses. There is room for objectivity in using the radiation wavelengths for colors, and pounds for masses.

Indeed, remember Rosetta and its adventure in space<sup>1</sup>: it took eleven years to *objectively* reach the 67P-Churyumov-Gerasimenko comet, and to *objectively* land the probe Philae on it. It was all based on an objective understanding of the law of

---

<sup>1</sup>See e.g. <https://www.nasa.gov/rosetta>.

gravitation, as well as on a variety of instruments developed thanks to our understanding of other physical laws. There must be some sense in which measurements are indeed objective! This is at once a goal and a fact, as beautifully expressed by Alexandre Koyré, while illustrating his view of Newton's science.<sup>2</sup>

To make sense of these ideas, in practice, let us observe that our measurement tools show at once only a given facet of the phenomenon at hand. Imagine that we are looking at the size of the orbitals of one atom of an aluminum bar. We don't see the bar, and we are not able to measure its length. We must zoom out and lose resolution for that; almost paradoxically, too much accuracy makes impossible the measurement of the length of the bar: on that scale the length of the bar has no physical relevance. Indeed, was this not the case, wouldn't we have liked to increase further the measurement accuracy, reaching the internal structure of the nuclei of the atoms? And then the internal structure of nucleons, and so on? As a matter of fact, this project is physically misguided: higher measurement accuracy requires more energy; but if the energy becomes too large, the object of interest may be destroyed, and its properties turn meaningless; what is (physically) the length of the bar, when it is vaporised?

Extreme resolutions may not be possible, not even in principle,<sup>3</sup> in fact they may even be disturbing rather than implying higher intelligibility.

Thus, besides the choice of observable quantities, tools and protocols, also the resolution, the scale of observation, is fundamental and must be properly chosen by the physicist. Once the details of what one wants to measure are set, different measurement methods lead to the same result *within the relevant degree of "accuracy"*. For instance, one may use different kinds of thermometers to measure the thermodynamic temperature of a gravitational bar, including electronic devices that look at microscopic vibrations and contact mercury thermometers. The measurement is accurate when the results are the same up to a certain number of digits, keeping in mind that many more digits would not correspond to the thermodynamic temperature, but perhaps to some other physical property.

---

<sup>2</sup>Citing Koyré, Ref. [2]: *Thus it seems to me that I have the right to assume that when we are speaking about Newton and Newtonianism we know more or less what we are speaking of. More or less! Somehow this very expression used in connection with Newton strikes me as improper, because it is possible that the deepest meaning and aim of Newtonianism, or rather, of the whole scientific revolution of the seventeenth century, of which Newton is the heir and the highest expression, is just to abolish the world of the "more or less," the world of qualities and sense perception, the world of appreciation of our daily life, and to replace it by the (Archimedean) universe of precision, of exact measures, of strict determination. To which Koyré adds: this characterization is very nearly equivalent to the mathematization (geometrization) of nature and therefore the mathematization (geometrization) of science. And finally: This, in turn, implies the disappearance—or the violent expulsion—from scientific thought of all considerations based on value, perfection, harmony, meaning, and aim, because these concepts, from now on merely subjective, cannot have a place in the new ontology.*

<sup>3</sup>For instance, chaotic dynamics increase uncertainties at an exponential rate, therefore accurate predictions on even relatively simple phenomena would rapidly need an accuracy on initial conditions requiring more energy than the object of interest can take.

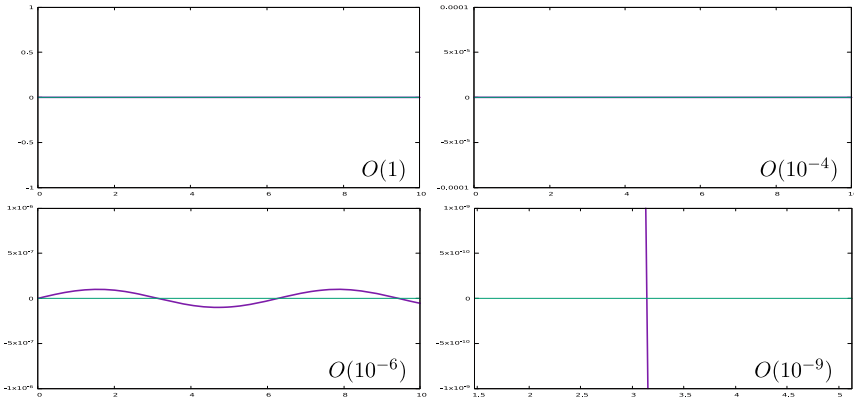
After the observables of interest have been identified, and the range of measurements has been delimited, a theory will connect the sets of measurements performed within those scales, hopefully constituting a satisfactory explanation of the observed connections between sets of data concerning different observables. If the scales are changed, hence the tools and the protocols, and perhaps also the observables are changed, a new theory will be needed. This new theory could be found to match with the previous one (finer or coarser than it may be) at the border between the two, or could be contained or contain the previous one. Often, however, the different theories do not mathematically reduce to each other [3], extra assumptions, alien to a given theory, are required to make the new theory mathematically agree with the old. After all, the same happens with our senses: a cathedral is meaningfully described with a language that differs from the one we might use to describe its stones; a movie is more meaningfully described in words that tell a story, than with the frequencies of the colors of the pixels of all its frames. It is not impossible or wrong to adopt a picture based on stones and pixels: the cathedral is made of stones, and the movie is made of pixels. But so much detail overshadows what we intend as the beauty or the meaning of the phenomenon of interest. In other words, our (present) knowledge of stones does not immediately reveal what makes a cathedral beautiful; what we (presently) know about pixels does not directly convey the message of a movie. Analogously, the (present) mathematical structures with which we describe atomic dynamics do not immediately lead to the second law of thermodynamics.

A fundamental point to understand is that a given theory is successful if the range of scales to which it properly applies is not too narrow, so that different people, bound to perform measurements in different ways, can still get compatible results. It will not be required that an agreement is found on all possible scales, indeed it will even be acceptable that on sufficiently separate scales, results are in some sense “orthogonal” to each other. We can illustrate that with the function

$$f(x) = 10^{-7} \sin x, \quad x \in [0; 2\pi]$$

assuming it represents a certain measurable quantity. Figure 1.1 shows that within a wide range of observations  $f$  is correctly represented by a straight horizontal line, it looks wavy within certain scales, and eventually it appears as a set of several vertical lines on still other scales. All these representations of  $f$  are legitimate and objective; the one that is also useful, or meaningful, depends on the scale we want to explore.

Truly, the choice of space and time scales, that of the observables of interest, and the choice of measurement tools and protocols are fundamental for the development of a physical theory. But this does not imply any subjectivity in the corresponding results. It should also be clear now that understanding better a given phenomenon does not mean to increase the resolution of measurements, in order to gain ever more detailed information. One should rather discard a large fraction of information, in order to highlight the *relevant* information. Galileo used to say: “*difalcare gli impedimenti*”, i.e. get rid of inessential details, in order to highlight the pure phenomenon of interest. Clearly, the message of a movie is not given by an accurate analysis of the pixels and sound bits of all frames of the film: it emerges when the frames pass before



**Fig. 1.1** Plots of  $f(x) = 10^{-7} \sin x$  at various resolutions  $O(10^{-k})$ ,  $k = 0, -4, -6, -9$ . For many levels of description,  $f$  can be treated as a straight horizontal line; for other resolutions it can be treated as several vertical straight lines (only one of which is reported). Only at intermediate resolutions does  $f$  look like a sinusoidal function. All representations are correct, but their usefulness depends on the scale of interest

our eyes at a speed that allows us to grasp just a tiny fraction of the information in them. Selecting what is essential and what is not, however, may be a hard task.

Which are the variables and the scales described by Thermodynamics? What is relevant to it? What is not relevant? What happens when scales are changed? Investigating these questions, which are particularly interesting for systems that are not of thermodynamic interest, is the subject of the present paper.

## 1.2 Continuum Description

Consider the simplest of the non-equilibrium thermodynamic relations, the mass continuity equation. This relation is derived considering mass as a continuum, which moves flowing through space, and one computes how much mass occupies a given volume in space by knowing how much mass is in it, how much flows in, and how much flows out, cf. Fig.1.2. Indeed, at sufficiently low energies that nuclear reactions are negligible, mass is a locally conserved quantity.<sup>4</sup>

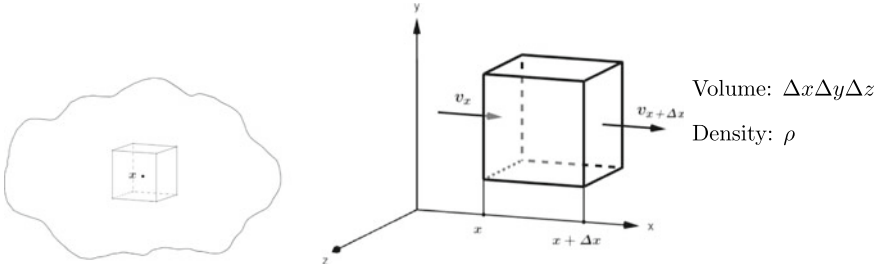
The mass entering per unit time from direction  $x$  is

$$\rho v_x \Big|_x \Delta y \Delta z \tag{1.1}$$

while that exiting is

$$\rho v_x \Big|_{x+\Delta x} \Delta y \Delta z \tag{1.2}$$

<sup>4</sup>As discussed in the Prologue, this statement makes sense only once the scale of interest has been specified, because nuclear reactions cannot be completely excluded.



**Fig. 1.2** Small cube containing a certain amount of mass  $\rho \Delta x \Delta y \Delta z$ , where  $\rho$  is the mass density and  $\Delta x, \Delta y, \Delta z$  its sides. This amount changes because mass flows in and out of the volume

The net variation per unit time due to motion in the  $x$  direction is thus given by:

$$\left[ \rho v_x \Big|_x - \rho v_x \Big|_{x+\Delta x} \right] \Delta y \Delta z \tag{1.3}$$

Analogously for directions  $y$  and  $z$ , one has:

$$\left[ \rho v_y \Big|_y - \rho v_y \Big|_{y+\Delta y} \right] \Delta x \Delta z \tag{1.4}$$

$$\left[ \rho v_z \Big|_z - \rho v_z \Big|_{z+\Delta z} \right] \Delta x \Delta y \tag{1.5}$$

The total content of mass in the volume is

$$\rho \Delta x \Delta y \Delta z \tag{1.6}$$

so the variation in a time  $\Delta t$  is

$$\left[ \rho \Big|_{t+\Delta t} - \rho \Big|_t \right] \Delta x \Delta y \Delta z \tag{1.7}$$

Then, the variation per unit time can be written as:

$$\begin{aligned} \frac{\rho \Big|_{t+\Delta t} - \rho \Big|_t}{\Delta t} \Delta x \Delta y \Delta z &= \left[ \rho v_x \Big|_x - \rho v_x \Big|_{x+\Delta x} \right] \Delta y \Delta z \\ &+ \left[ \rho v_y \Big|_y - \rho v_y \Big|_{y+\Delta y} \right] \Delta x \Delta z \\ &+ \left[ \rho v_z \Big|_z - \rho v_z \Big|_{z+\Delta z} \right] \Delta x \Delta y \end{aligned} \tag{1.8}$$

Divide by  $\Delta x \Delta y \Delta z$ :

$$\frac{\rho|_{t+\Delta t} - \rho|_t}{\Delta t} = \frac{\rho v_x|_x - \rho v_x|_{x+\Delta x}}{\Delta x} + \frac{\rho v_y|_y - \rho v_y|_{y+\Delta y}}{\Delta y} + \frac{\rho v_z|_z - \rho v_z|_{z+\Delta z}}{\Delta z} \quad (1.9)$$

In the  $\Delta t \rightarrow 0$ ,  $\Delta x \rightarrow 0$ ,  $\Delta y \rightarrow 0$ ,  $\Delta z \rightarrow 0$  limit, we obtain:

$$\frac{\partial \rho}{\partial t} = -\text{div}(\rho v) \quad \text{Continuity Equation} \quad (1.10)$$

What have we assumed? That mass is conserved and that mass and velocity fields are not only continuous, but *differentiable* quantities. If matter is made of *atoms*, we have however a problem. In fact, the density at a point  $x$  is defined as the ratio of the mass  $m$  to the volume  $V$  containing it, around a point  $x$ . The ratio is a function of the volume, hence the following limit may be considered, to make it a robust property of the system at  $x$ :

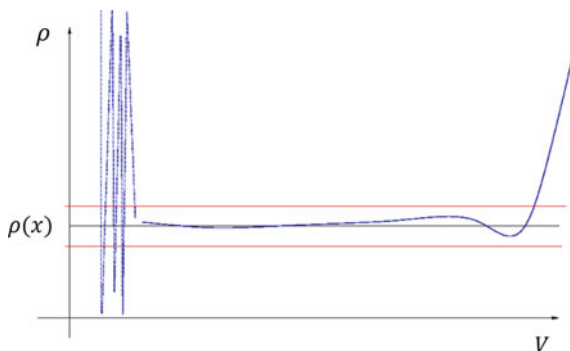
$$\rho(x) = \lim_{V \rightarrow 0} \frac{m(V)}{V} \Big|_x \quad (1.11)$$

Of course, limit does not mean that  $V$  reaches 0, but that given a certain tolerance  $\epsilon > 0$  on the measurements of  $\rho$ , there is a volume  $\delta > 0$  such that the ratio of mass and volume changes less than  $\epsilon$ , if the volume changes within  $(0, \delta)$ . Therefore,  $\delta$  can be regarded as the accuracy of volume measurements required to meet the desired accuracy in measurements of the density at  $x$ . If matter is a continuum, we cannot and do not need to qualify more precisely  $\epsilon$  and  $\delta$ ; but the fact that matter is made of atoms imposes certain constraints on both. In particular, it makes no sense to measure the density on scales of the atomic size or smaller: that scale pertains to atomic physics, not to thermodynamics. On the atomic scale, the macroscopic notion of mass density, as a continuum, makes no useful sense. The picture may be something like Fig. 1.3.

Does this mean that thermodynamics is just a *coarse* or *subjective* description? It depends. It is meaningful and objective when the macroscopic scale is widely separated from the microscopic scale, so that measurements performed within many orders of magnitude of  $V$  all yield the same density, i.e. they are indistinguishable and make all observers agree within that level of reality. There may be a range of inaccurate measurements (large volumes), within which the density changes substantially when the accuracy is improved (reducing  $V$ ). But this can be followed by a wide range of  $V$  for which the measurement of the density remains within narrow bounds (red lines), hence it takes a precise, objective value. Finally, there will be a range within which the notion of density is meaningless (atomic scale).

Similar reasonings apply to other conserved and non-conserved (having sources and sinks) quantities, such as momentum, energy, entropy etc. The question is: How do we know that matter is made of atoms?





**Fig. 1.3** Hypothetical density versus volume graph. For large volumes, the density changes in a macroscopic measurable fashion, due to the inhomogeneities of the mass distribution in space: the notion of density makes sense but the measurement is not accurate. When volume decreases below a certain threshold and for a wide range of scales, the density only varies within a narrow band of values: measurement is accurate; density is a sensible objective quantity. At atomic scales and smaller, the ratio mass to volume wildly fluctuates, depending on whether the volume contains or does not contain atoms or parts of atoms. There is nothing wrong with this fact, but the measurement of such a frantically fluctuating quantity is no use: density has lost significance as a quantity that can be described by an intelligible equation, such as the continuity equation (1.10)

### 1.3 Brownian Motion: Fluctuations Reveal Atoms

Suppose a particle of mass  $m$  is dragged by a constant force  $F$  in a fluid, that is a continuous viscous medium. Letting  $x$  be the position of the particle and  $\dot{x} = v$  its velocity, its motion is described by the following equations:

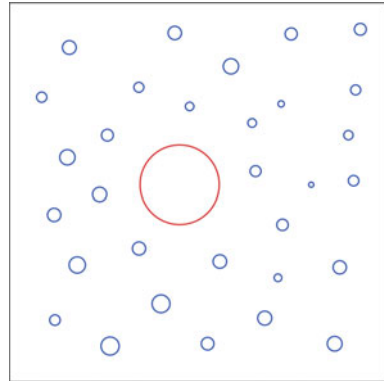
$$m\dot{v} + \alpha v = F; \quad v(t) = Ce^{-t/\tau} + v_\infty; \quad v_\infty = \frac{1}{\alpha}F = \frac{1}{m\gamma}F; \quad \mu = \frac{1}{m\gamma} \quad (1.12)$$

Here,  $\gamma = \alpha/m$ ,  $\tau = 1/\gamma$  is the characterisitic time scale of the phenomenon,  $\mu$  is the *mobility* of that particle in that fluid, i.e. the proportionality constant between the asymptotic velocity and the driving force, and the friction force is expressed by Stoke’s law  $F_c = -\alpha v$ , where the dimensions of  $\alpha$  are  $[\alpha] = kg/s$ . If there is no force dragging the particle,  $F = 0$ , the solution of (1.12) is:

$$v(t) = v(0)e^{-t/\tau} \quad (1.13)$$

As an example, let the particle be spherical of radius  $a$ , so that the friction coefficient is given by  $\alpha = 6\pi a\eta$ , where  $\eta$  is the fluid viscosity. Let the little sphere represent a pollen grain, then, typically  $a = O(10^{-2})$  cm and  $m = O(10^{-7})$  g. Supposing the fluid is water at room temperature and pressure, one has  $\eta_{H_2O} = 10^{-2}g/(cms)$ . Hence,  $\tau = O(10^{-4})$  s, and whatever the initial speed, this equation predicts that in one second  $v(0)$  should be reduced by a factor  $\exp(-10^4)$ !

**Fig. 1.4** A large Brownian particle is suspended in a liquid made of much smaller particles. They all follow the laws of classical mechanics, conserving energy and momentum at collisions



In contrast to that, one observes that the particle continues to move, and does not reveal any trend towards immobility; rather the particle moves erratically, in a totally unpredictable fashion. Apparently, Stokes' force is not the only one acting on the particle, and because no external action is applied, there must be forces that are generated by the fluid, in addition to the viscous force. Moreover, these forces do not seem to play a role when macroscopic objects are immersed in the fluid. Therefore, very light objects such as pollen grains, might be affected by impacts with even lighter objects: the molecules that constitute the fluid; that would explain the randomness of the motion of pollen. In fact, one of the first to propose this picture was Cantoni, who made an impressive series of experiments, in which he also gave evidence of equipartition of energy, and eventually concluded [4]:

I think that the dancing movement of the extremely minute solid particles in a liquid, can be attributed to the different velocities that must be proper at a given temperature of both such solid particles and of the molecules of the liquid that hit them from every side. I do not know whether others did already attempt this way of explaining Brownian motions

Still, it was not clear why molecules should exert at once the viscous force that characterizes continua, and the conservative force arising in collisions with the object of interest; it was not clear how tiny invisible objects could affect small but visible, hence much larger, objects<sup>5</sup> and the effect could have even been contrary to thermodynamics.

Einstein, and independently Smoluchowski, took up the task of explaining the Brownian motion, accepting this scenario. In particular, Einstein remarked that particles, even if very large, must obey the laws of statistical mechanics and, in particular generate osmotic pressure, just as with ordinary solutions. The description of the motion of suspended particles could thus be formulated as depicted in Fig.1.4.

There are particles of mass  $m_i$  constituting the fluid, and a large particle of mass  $m$  suspended in it, which obey the laws of classical mechanics. Their trajectories can be computed, once the initial conditions are given, by integrating the equations of motion:

---

<sup>5</sup>The mass of a water molecule is about  $10^{16}$  times smaller than that of pollen grains.

$$\dot{x} = v, \quad m\dot{v} = -\alpha v + \sum_{i=1}^N F_i; \quad \dot{x}_i = v_i, \quad m_i\dot{v}_i = -F_i + \sum_{\substack{j=1 \\ j \neq i}}^N F_{ji} \quad (1.14)$$

where  $F_i$  represents the interaction of the  $i$ th particle with the mass  $m$ ; and  $F_{ji}$  the interaction among the  $i$ -th and  $j$ th particles. Supposing there is a mole of water molecules, the indices  $i$  and  $j$  run from 1 to Avogadro's number  $N_A$ , which at the time was not known. This picture places the motion of pollen in an intermediate situation, between the microscopic and the macroscopic dynamics, being well distinguished from the two, because the microscopic motions do not contemplate viscosity, and the macroscopic description does not account for molecular impacts. Denoting by  $m\Gamma$  the force on the large particle resulting from all molecular impacts, one may then write:

$$\dot{v} = -\gamma v + \Gamma \quad (1.15)$$

with initial conditions

$$x(0) = x_0, \quad v(0) = v_0, \quad x_1(0) = x_{10}, \quad v_1(0) = v_{10}, \dots, \quad x_N(0) = x_{N0}, \quad v_N(0) = v_{N0} \quad (1.16)$$

Integration yields:

$$v(t) = e^{-\gamma t} \left( v_0 + \int_0^t e^{\gamma s} \Gamma(s; x_0, \dots, v_{N0}) ds \right) \quad (1.17)$$

where we stress that the force  $m\Gamma$  depends on the initial conditions and on time. This direct approach meets, however, insurmountable problems, because one cannot solve  $10^{23}$  equations of motion in practice, and also because the initial conditions of the water molecules are not known. Furthermore, even if all necessary information were available, and equations could actually be solved, the approach would be useless. Every time the process is repeated, the initial condition is different, the calculations would have to be repeated, and no prediction would ever been made: knowledge of previous calculations/experiments would be useless for future experiments; no real understanding of the phenomenon could be claimed.

A different approach is necessary. The idea was to pass from the exact description of the phenomenon to a statistical description, which is less detailed, but more meaningful, cf. Sect. 1.1. In practice, one may consider a large number of independent suspended particles or, equivalently, a sequence of experiments performed with a single particle, so that the possible initial conditions are adequately sampled, and then one may be satisfied with the corresponding averages. This way, the equations of motion do not need precise knowledge of the force  $m\Gamma$ ; its statistical properties, that can be inferred from observation, will do. The result will not allow us to make predictions on the single suspended particle, unless some special condition is veri-

fied,<sup>6</sup> but that may indeed be of no interest. Let us then do as if  $\Gamma$  were not known, apart from its average and its two times autocorrelation function. Observation suggests the following:

- in an isolated system, the forces  $F_i$  do no net work on a mass  $m$ , otherwise they would use internal energy of the fluid, that would be eventually exhausted, leading the process to halt. In a closed system,<sup>7</sup> heat could flow to the system to keep it going, but then net work would result in a drift of the center of mass of the suspended particles. This is not observed in experiments, therefore one may write:

$$\mathbb{E}[\Gamma(t)] = 0 \quad (1.18)$$

where  $\mathbb{E}[\Gamma(t)]$  stands for the expectation (mean) value of  $\Gamma$  at any time  $t$ , which is computed averaging over the collection of particles or of experiments under consideration.

- the motion of a suspended particle appears uncorrelated in time if observed at times  $t$  and  $t'$  separated by sufficiently large time intervals. Therefore, the accelerations due to the unknown forces should satisfy:

$$\mathbb{E}[\Gamma(t)\Gamma(t')] = 0 \quad \text{for } |t - t'| \geq \tau_0 > 0 \quad (1.19)$$

where the correlation time  $\tau_0$  is empirically determined for the fluid and the suspended particles at hand;

- in the Brownian motion,  $\tau_0$  appears even shorter than  $\tau = 1/\gamma$ , therefore for sake of simplicity, one may write:

$$\mathbb{E}[\Gamma(t)\Gamma(t')] = q\delta(t - t') \quad (1.20)$$

where  $\delta$  is the Dirac delta function and  $q$  an unspecified constant.

Because, the force  $m\Gamma$  is known only on average, we cannot limit its impact on the velocity of  $m$ . Therefore, we cannot guarantee the differentiability of the process, and the equation of motion is better expressed by something like:

$$dv + \gamma v dt = d\tilde{\Gamma} \quad (1.21)$$

where the differentials  $dv$  and  $d\tilde{\Gamma}$  represent the variations of the functions  $v$  and  $\tilde{\Gamma}$  in an elementary time interval  $dt$ . This way one indicates that such functions are

---

<sup>6</sup>For instance, it could happen that all the different values of a given microscopic quantity are experienced by a single particle in time. Then, averaging over the ensemble of independent particles may yield the average of the values experienced in time by the single particle. Whether this actually happens or not depends on the system and on the quantity at hand, and only experience can tell. In any case, this requires observation scales substantially longer than the scales of the microscopic events.

<sup>7</sup>A closed system exchanges energy with the environment, but no mass.

not bound to be differentiable in time. Only when the operation  $\Gamma = d\tilde{\Gamma}/dt$  makes sense, can one define the derivative of  $v$  with respect to time and use (1.15).

Equation (1.21) is called *stochastic differential equation*. One may formally integrate it, but it yields concrete results only on the statistical properties of the quantities appearing in it. Therefore, its terms should not be interpreted literally as velocities or forces; in particular, one should expect  $m\Gamma$  to obey the action-reaction principle, especially in nonequilibrium conditions, see e.g. [5–8]. Let us integrate it first, with formal initial condition  $v(0) = v_0$ . One obtains:

$$v(t) = e^{-\gamma t} \left( v_0 + \int_0^t e^{\gamma t'} \Gamma(t') dt' \right) = v_0 e^{-\gamma t} + \int_0^t e^{-\gamma(t-t')} \Gamma(t') dt' \quad (1.22)$$

Then, averaging, one can write:

$$\mathbb{E}[v(t)] = v_0 e^{-\gamma t} + \int_0^t e^{-\gamma(t-t')} \mathbb{E}[\Gamma(t')] dt' = v_0 e^{-\gamma t} \quad (1.23)$$

because  $\mathbb{E}[\Gamma(t')] = 0$ . This result means that *on average* the velocity of suspended (Brownian) particles follows the classical hydrodynamic law: although no particle slows down, *on average* their speed decreases exponentially in time.<sup>8</sup> One may then compute the correlation function, starting from the following formal expression:

$$\begin{aligned} v(t_1)v(t_2) &= \left( v_0 e^{-\gamma t_1} + \int_0^{t_1} e^{-\gamma(t_1-t'_1)} \Gamma(t'_1) dt'_1 \right) \left( v_0 e^{-\gamma t_2} + \int_0^{t_2} e^{-\gamma(t_2-t'_2)} \Gamma(t'_2) dt'_2 \right) \\ &= v_0^2 e^{-\gamma(t_1+t_2)} + v_0 \int_0^{t_2} e^{-\gamma(t_1+t_2-t'_2)} \Gamma(t'_2) dt'_2 + v_0 \int_0^{t_1} e^{-\gamma(t_1+t_2-t'_1)} \Gamma(t'_1) dt'_1 \\ &\quad + \int_0^{t_1} dt'_1 \int_0^{t_2} dt'_2 e^{-\gamma(t_1+t_2-t'_1-t'_2)} \Gamma(t'_1) \Gamma(t'_2) \end{aligned} \quad (1.24)$$

Averaging again over all realizations of the process, we obtain

$$\mathbb{E}[v(t_1)v(t_2)] = v_0^2 e^{-\gamma(t_1+t_2)} + v_0 \int_0^{t_1} e^{-\gamma(t_1+t_2-t'_1)} \mathbb{E}[\Gamma(t'_1)] dt'_1 + v_0 \int_0^{t_2} e^{-\gamma(t_1+t_2-t'_2)} \mathbb{E}[\Gamma(t'_2)] dt'_2$$

---

<sup>8</sup>Note that the averaging operation has been exchanged with the time integral in (1.23). This is justified, since one averages over a large but always finite number of independent particles, hence  $\mathbb{E}[v]$  can be understood as  $(1/N) \sum_{i=1}^N v_i$ , where  $i$  denotes the  $i$ th Brownian particle.

$$\begin{aligned}
& + \int_0^{t_1} dt'_1 \int_0^{t_2} dt'_2 e^{-\gamma(t_1+t_2-t'_1-t'_2)} \mathbb{E}[\Gamma(t'_1)\Gamma(t'_2)] \\
& = v_0^2 e^{-\gamma(t_1+t_2)} + \int_0^{t_1} dt'_1 \int_0^{t_2} dt'_2 e^{-\gamma(t_1+t_2-t'_1-t'_2)} q \delta(t'_1 - t'_2)
\end{aligned} \tag{1.25}$$

To compute the double integral, we consider two cases:  $t_1 > t_2$  and  $t_1 < t_2$ . For  $t_1 > t_2$ , we can write:

$$\begin{aligned}
& \int_0^{t_2} dt'_2 \int_0^{t_1} dt'_1 e^{-\gamma(t_1+t_2-t'_1-t'_2)} q \delta(t'_1 - t'_2) = q e^{-\gamma(t_1+t_2)} \int_0^{t_2} e^{2\gamma t'_2} dt'_2 \\
& = \frac{q}{2\gamma} [e^{-\gamma(t_1-t_2)} - e^{-\gamma(t_1+t_2)}] = \frac{q}{2\gamma} [e^{-\gamma|t_1-t_2|} - e^{-\gamma(t_1+t_2)}]
\end{aligned} \tag{1.26}$$

and swapping  $t_1$  and  $t_2$ , for  $t_1 < t_2$  we have:

$$\frac{q}{2\gamma} [e^{-\gamma(t_2-t_1)} - e^{-\gamma(t_1+t_2)}] = \frac{q}{2\gamma} [e^{-\gamma|t_1-t_2|} - e^{-\gamma(t_1+t_2)}] \tag{1.27}$$

Finally, combining the two cases, yields:

$$\mathbb{E}[v(t_1)v(t_2)] = \frac{q}{2\gamma} [e^{-\gamma|t_1-t_2|} - e^{-\gamma(t_1+t_2)}] + v_0^2 e^{-\gamma(t_1+t_2)} \tag{1.28}$$

which holds for  $t_1 = t_2 = t$  as well.<sup>9</sup> Then, for large  $t_1$  and  $t_2$ ,  $\mathbb{E}[v(t_1)v(t_2)]$  does not depend on the initial velocity  $v_0$  and takes the form:

$$\mathbb{E}[v(t_1)v(t_2)] \approx \frac{q}{2\gamma} e^{-\gamma|t_1-t_2|} \tag{1.29}$$

and the asymptotic average of the kinetic energy  $E$  writes:

$$\mathbb{E}[E] = \lim_{t \rightarrow 0} \frac{1}{2} m \mathbb{E}[v(t)v(t)] = \frac{mq}{4\gamma} \tag{1.30}$$

If we assume that the system is in equilibrium (no external forces, no dissipation) and energy follows the equipartition principle,  $\mathbb{E}[E] = k_B T/2$ , we obtain:

$$\frac{1}{2} k_B T = \frac{mq}{4\gamma} \Rightarrow q = \frac{2k_B T \gamma}{m} \tag{1.31}$$

---

<sup>9</sup>It is only a matter of recalling the physical meaning of the Dirac  $\delta$ : that of a very peaked and correspondingly narrow function.

Consider now the squared distance travelled by a particle in a time  $t$ :

$$(x(t) - x_0)^2 = \int_0^t v(t_1) dt_1 \int_0^t v(t_2) dt_2 = \int_0^t \int_0^t v(t_1)v(t_2) dt_1 dt_2 \quad (1.32)$$

Again, averaging over all realizations, one obtains

$$\begin{aligned} \mathbb{E}[(x(t) - x_0)^2] &= \int_0^t \int_0^t \mathbb{E}[v(t_1)v(t_2)] dt_1 dt_2 \\ &= \int_0^t dt_1 \int_0^t dt_2 \left[ \frac{q}{2\gamma} (e^{-\gamma|t_1-t_2|} - e^{-\gamma(t_1+t_2)}) + v_0^2 e^{-\gamma(t_1+t_2)} \right] \\ &= \int_0^t dt_1 \int_0^t dt_2 \left[ \frac{q}{2\gamma} e^{-\gamma|t_1-t_2|} + \left( v_0^2 - \frac{q}{2\gamma} \right) e^{-\gamma(t_1+t_2)} \right] \end{aligned} \quad (1.33)$$

where

$$\int_0^t \int_0^t e^{-\gamma(t_1+t_2)} dt_1 dt_2 = \int_0^t e^{-\gamma t_1} dt_1 \int_0^t e^{-\gamma t_2} dt_2 = \frac{1}{\gamma^2} (e^{-\gamma t} - 1)^2 \quad (1.34)$$

and

$$\begin{aligned} \int_0^t \int_0^t e^{-\gamma|t_1-t_2|} dt_1 dt_2 &= \int_0^t \left[ \int_0^{t_1} e^{-\gamma(t_1-t_2)} dt_2 + \int_{t_1}^t e^{\gamma(t_1-t_2)} dt_2 \right] dt_1 \\ &= \int_0^t \left[ e^{-\gamma t_1} \frac{e^{\gamma t_1} - 1}{\gamma} + e^{\gamma t_1} \frac{e^{-\gamma t_1} - e^{-\gamma t}}{\gamma} \right] dt_1 \\ &= \int_0^t \left[ \frac{2}{\gamma} - \frac{1}{\gamma} e^{-\gamma t_1} - \frac{1}{\gamma} e^{\gamma t_1} e^{-\gamma t} \right] dt_1 \\ &= \frac{2}{\gamma} t + \frac{1}{\gamma^2} (e^{-\gamma t} - 1) - \frac{1}{\gamma^2} e^{-\gamma t} (e^{\gamma t} - 1) \\ &= \frac{2}{\gamma} t + \frac{2}{\gamma^2} (e^{-\gamma t} - 1) \end{aligned} \quad (1.35)$$

Eventually, one can write:

$$\mathbb{E}[(x(t) - x_0)^2] = \left( v_0^2 - \frac{q}{2\gamma} \right) \frac{(1 - e^{-\gamma t})^2}{\gamma^2} + \frac{q}{2\gamma} \left[ \frac{2}{\gamma} t - \frac{2(1 - e^{-\gamma t})}{\gamma^2} \right] \quad (1.36)$$

which explodes linearly in time. One may then introduce the diffusion coefficient as:

$$D := \lim_{t \rightarrow \infty} \frac{\mathbb{E}[(x(t) - x_0)^2]}{2t} = \frac{1}{2} \lim_{t \rightarrow \infty} \left[ \frac{q}{\gamma^2} + \frac{v_0^2 - \frac{q}{2\gamma}}{\gamma^2 t} - \frac{q}{\gamma^3 t} \right] = \frac{q}{2\gamma^2} = \frac{k_B T}{m\gamma} \quad (1.37)$$

where the last equality holds if equipartition does. Equation (1.37) yields the celebrated Einstein relation:

$$D = \mu k_B T \quad (1.38)$$

which links the diffusion coefficient to the temperature via the mobility.

The presence of fluctuations, i.e. the mean square displacement  $\mathbb{E}[(x(t) - x_0)^2]$ , and of dissipation,  $\mu$ , makes Eq. (1.37) the first Fluctuation-Dissipation Relation. It is an amazing result experimentally validated first by Perrin [9], that allows the calculation of Avogadro's number  $N_A$ . Indeed, it is relatively simple to measure the variance of positions, which yields  $D$ ; the temperature  $T$ , the friction coefficient  $\gamma$  and the mass  $m$  are known, and the Boltzmann constant can be expressed as the ratio  $k_B = R/N_A$ , where  $R$ , the universal constant of gases, is also known. Then one obtains:

$$D = \frac{k_B T}{m\gamma} = \frac{RT}{6\pi\eta a} \frac{1}{N_A} \quad (1.39)$$

where the only unknown is  $N_A$ , and can thus be estimated to be  $N_A = 6.02210^{23} \text{ mol}^{-1}$ . If atoms can be counted, they definitely exist!

The above is readily generalized to a system in a  $d$ -dimensional space, for which one has:

$$\dot{v}_i = -\gamma v_i + \Gamma_i(t), \quad i = 1, \dots, d \quad (1.40)$$

In case there are no correlations among the different coordinates, one may assume:

$$\mathbb{E}[\Gamma_i(t)] = 0; \quad \mathbb{E}[\Gamma_i(t)\Gamma_j(t')] = q\delta_{ij}\delta(t - t') \quad (1.41)$$

so that each variable can be treated separately. Moreover, when equipartition applies, we can write:

$$\mathbb{E}[E] = \sum_{i=1}^d \frac{1}{2} m \mathbb{E}[v_i^2] = \frac{d}{2} k_B T, \quad \text{and} \quad \frac{\mathbb{E}[\overrightarrow{\Delta x}(t)^2]}{t} \rightarrow \frac{2dk_B T}{m\gamma}, \quad \text{for } t \rightarrow \infty \quad (1.42)$$

where  $\overrightarrow{\Delta x}(t)$  is  $(\mathbf{x}(t) - \mathbf{x}(0))$  and has dimension  $d$ .

Relation (1.39) contains in itself an incredible amount of information. In the first place, it links the microscopic realm with the macroscopic one through  $k_B =$



$1.380649 \cdot 10^{-23} \text{ J/K}$ , which marks the huge “distance” between the two. Then, once the value of  $N_A$  has been established and confirmed in many experiments, Eq. (1.39) allows us to compute a macroscopic nonequilibrium property, the viscosity  $\eta$ , from an equilibrium experiment in which no work is done: fluctuations are merely observed. Note: viscosity does not exert any force when the fluid is at rest with respect to its container or to the walls of an object in it. In a direct experimental measurement of the viscosity, the object in the fluid is dragged by a force  $F$ ; then, viscosity imposes a limit velocity  $v_\infty$ , and the mobility is defined by the linear response relation  $F = \mu v_\infty$ . Equation (1.39) bypasses all that work, and relates non-equilibrium properties to equilibrium fluctuations, which are made visible by objects belonging to an intermediate, mesoscopic, realm.

Hard to beat the beauty and importance of such a result, which settled once and for all the dispute about the existence of atoms, whose mechanics provides a microscopic description of the state of a macroscopic object. The theory of Brownian motion, experimentally demonstrated by Perrin [9] dispelled all doubts, and several decades later Feynman could state:

If, in some cataclysm, all of scientific knowledge had to be destroyed, and only one sentence passed on to the next generation of creatures, what statement would contain the most information in the fewest words? I believe it is the atomic hypothesis (or the atomic fact, or whatever you wish to call it) that all things are made of atoms—little particles that move around in perpetual motion, attracting each other when they are a little distance apart, but repelling upon being squeezed into one another. In that one sentence, you will see, there is an enormous amount of information about the world, if just a little imagination and thinking are applied.

Imagination is needed because even solving the corresponding equations of motion, something so far impossible, would not help us understand the macroscopic world. What would we do with myriads of graphs and numbers with very detailed information on atoms trajectories, if we are concerned, for instance, about warming up or cooling down something? These questions are much more naturally approached by a very useful macroscopic description, not resting at all on microscopic theories: Thermodynamics. However, there is catch: in order for the macroscopic approach to be useful, its laws must be supplemented not only with a specification of the appropriate boundary conditions but with the values of thermophysical constants such as the transport coefficients. These values cannot be predicted by the macroscopic theory, and must be supplied by experiments or derived from other approaches. One of the goals of statistical mechanics is to predict these parameters from knowledge of the interactions of the system’s constituent molecules.

### 1.3.1 Langevin Treatment

The solutions of the Langevin equation (1.15), or (1.21), define a stochastic process: a process that is intrinsically probabilistic, in which only the statistical properties of the variables of interest are relevant. There is an incredible variety of stochastic

processes. We call Gaussian a process  $Z(t)$  indexed by a parameter  $t$ , e.g. time, if the joint probability density of its values  $z_1, z_2, \dots, z_n$  at times  $t_1, t_2, \dots, t_n$  is given by

$$W_n(z_1, t_1; \dots; z_n, t_n) = ce^{-\frac{1}{2} \sum_{i,j=1}^n a_{ij}(z_i - m_i)(z_j - m_j)} \quad (1.43)$$

where  $m_k = \mathbb{E}[Z(t_k)]$  and  $A = (a_{ij})$  is a positive definite  $n \times n$  matrix, whose inverse has entries defined by  $(A^{-1})_{ij} = \mathbb{E}[(Z(t_i) - m_i)(Z(t_j) - m_j)]$ , for  $i, j = 1, \dots, n$ . The first three *cumulants* of the probability distribution of a random variable  $Z$ ,  $\mathbb{E}[\cdot]_c$ , are defined by:

$$\begin{aligned} \mathbb{E}[Z]_c &= \mathbb{E}[Z] \\ \mathbb{E}[Z^2]_c &= \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 \\ \mathbb{E}[Z^3]_c &= \mathbb{E}[Z^3] - 3\mathbb{E}[Z^2]\mathbb{E}[Z] + 2\mathbb{E}[Z]^3 \end{aligned} \quad (1.44)$$

Therefore, a normal distribution  $N(m, \sigma)$  implies:

$$\mathbb{E}[Z]_c = m, \quad \mathbb{E}[Z^2]_c = \sigma^2, \quad \mathbb{E}[Z^n]_c = 0 \quad \text{for } n \geq 3 \quad (1.45)$$

and if all cumulants of order  $n \geq 3$  vanish, then the distribution is normal. defining a linear transformation of a stochastic process  $Z(t)$  as

$$Y(t) = \int_a^b c(t, t')Z(t')dt' \quad (1.46)$$

the cumulant of  $Y(t)$ , for any order  $n$ , is given by:

$$\mathbb{E}[Y^n]_c = \int_a^b dt'_1 \int_a^b dt'_2 \dots \int_a^b dt'_n c(t_1, t'_1)c(t_2, t'_2) \dots c(t_n, t'_n)\mathbb{E}[Z^n]_c \quad (1.47)$$

Hence, if  $Z(t)$  is Gaussian,  $Y(t)$  is, because all cumulants of order  $n \geq 3$  vanish. Then, the velocity  $V$  of the Brownian motion is a Gaussian process if the stochastic term  $\Gamma(t)$  is, because it results from an integral like (1.46). Considering that in the  $t \rightarrow \infty$  limit,  $\mathbb{E}[V]_c = 0$ , and  $\sigma^2 = \mathbb{E}[V^2]_c = q/2\gamma$ , the asymptotic probability density of  $V$  is given by:

$$f_V(v) = ce^{-\frac{v^2}{(2q/2\gamma)}} = ce^{-\frac{\gamma v^2}{q}} = \sqrt{\frac{m}{2\pi k_B T}} e^{-\frac{mv^2}{2k_B T}} = f_{MB}(v) \quad (1.48)$$

if we accept the equipartition principle,  $\frac{q}{\gamma} = \frac{2k_B T}{m}$ .

This is the celebrated *Maxwell-Boltzmann* distribution in 1 dimension. Because its variance takes the form  $\sigma^2 = k_B T/m$ , fixing the temperature of the fluid  $T$ , and placing a large mass  $m$  in it, yields small uncertainty on the velocity: the velocity

is practically deterministic. On the contrary, a small mass  $m$  implies large uncertainty: in the small  $m$  limit, the velocity is totally random. This can already be seen in the Langevin equation, when the intensity of  $\Gamma$  is expressed by  $q$ : large  $q/\gamma$  (relatively small mass) means that random forces dominate the deterministic viscous force, while small  $q/\gamma$  is the opposite situation in which hydrodynamics holds. For a test particle of the size of water molecules, the motion is random and there is no viscosity, while for a boat it is deterministic, and atomic impacts are negligible. The Brownian particle lays at the border between the micro- and the macro-worlds that coexist, even though *very* different, in a single reality, and reveals both. While microscopic dynamics know no viscosity and no energy dissipation, which is why the motion does not stop, macroscopic dynamics knows no molecules, it dissipates energy, hence the motion stops. These two pictures are not contradictory: the ordered motion of the macroscopic object passes its energy to the disordered motion of the fluid molecules, contributing to the internal energy of the fluid; no energy is lost, it is simply randomized.

Let us introduce the notion of transition probability density  $P$ , to express the time evolution of a probability density  $W$ :

$$W(z, t + \tau) = \int P(z, t + \tau | z', t) W(z', t) dz' \quad (1.49)$$

which means that the probability density at time  $t + \tau$  that the random variable  $Z$  takes values around  $z$  cumulates the contributions coming from all possible values  $z'$  at time  $t$ , weighted with the probability density of transiting from  $z'$  at time  $t$ , to  $z$  at time  $t + \tau$ . Suppose the moments of  $P$  are known:

$$M_n(z', t, \tau) = \int_{-\infty}^{+\infty} (z - z')^n P(z, t + \tau | z', t) dz \quad (1.50)$$

one can construct the characteristic function, defined by:

$$c(u, z', t, \tau) = \int_{-\infty}^{+\infty} e^{iu(z-z')} P(z, t + \tau | z', t) dz = 1 + \sum_{n=1}^{+\infty} \frac{(iu)^n}{n!} M_n(z', t, \tau) \quad (1.51)$$

which can be inverted to give:

$$P(z, t + \tau | z', t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-iu(z-z')} c(u, z', t, \tau) du \quad (1.52)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-iu(z-z')} \left[ 1 + \sum_{n=1}^{+\infty} \frac{(iu)^n}{n!} M_n(z', t, \tau) \right] du \quad (1.53)$$

Introducing the following representation of  $\delta$  function:

$$\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk e^{ikx} \quad (1.54)$$

we get:

$$P(z, t + \tau | z', t) = \left[ 1 + \sum_{n=1}^{+\infty} \frac{1}{n!} \left( -\frac{\partial}{\partial z} \right)^n M_n(z, t, \tau) \right] \delta(z - z') \quad (1.55)$$

which yields:

$$W(z, t + \tau) = \int W(z', t) \delta(z - z') dz' - \frac{\partial}{\partial z} M_1(z, t, \tau) \int W(z', t) \delta(z - z') dz' + \dots \quad (1.56)$$

$$= \left[ 1 - \frac{\partial}{\partial z} M_1 + \frac{1}{2} \frac{\partial^2}{\partial z^2} M_2 + \dots \right] W(z, t) \quad (1.57)$$

This expression is useful to derive an evolution equation for  $W$ . Taking to the left hand side the term  $W(z, t)$ , dividing by  $\tau$  and taking the  $\tau \rightarrow 0$  limit, one obtains:

$$\lim_{\tau \rightarrow 0} \frac{W(z, t + \tau) - W(z, t)}{\tau} = \frac{\partial W}{\partial t} = \lim_{\tau \rightarrow 0} \sum_{n=1}^{+\infty} \frac{1}{n!} \left( -\frac{\partial}{\partial z} \right)^n \frac{M_n(z, t, \tau)}{\tau} W(z, t) \quad (1.58)$$

which is the Kramers-Moyal expansion.

For a process  $Z(t)$  obeying the Langevin equation with Gaussian,  $\delta$ -correlated noise, the moments of order higher than 2 vanish:  $M_n = 0$  for  $n \geq 3$ . For  $n = 1, 2$ , consider the following equation:

$$\dot{z} = h(z, t) + g(z, t)\Gamma(t), \quad \mathbb{E}[\Gamma(t)] = 0; \quad \mathbb{E}[\Gamma(t)\Gamma(t')] = 2\delta(t - t') \quad (1.59)$$

One obtains [10]:

$$\begin{aligned} D^{(1)}(z, t) &= h(z, t) + g'(z, t)g(z, t) \\ D^{(2)}(z, t) &= g^2(z, t) \end{aligned} \quad \text{where } D^{(k)}(z, t) = \lim_{\tau \rightarrow 0} \frac{1}{k!} \frac{M_k(z, t, \tau)}{\tau} \quad (1.60)$$

The term  $D^{(1)}$  is called drift coefficient, and  $D^{(2)}$  is called diffusion coefficient because they appear in the Fokker-Planck equation:

$$\frac{\partial W}{\partial t} = \left[ -\frac{\partial}{\partial z} D^{(1)}(z, t) + \frac{\partial^2}{\partial z^2} D^{(2)}(z, t) \right] W(z, t) = -\frac{\partial}{\partial z} J(z, t) \quad (1.61)$$

formally as contributions from a probability drift and from the diffusion of probability, with

$$J(z, t) = D^{(1)}(z, t)W(z, t) - \frac{\partial}{\partial z} D^{(2)}(z, t)W(z, t) \tag{1.62}$$

representing a probability current. The Fokker-Planck equation is thus a balance equation, analogous to those of conserved quantities in thermodynamics, since probability is conserved.

For more general cases, in which the number of stochastic variables is  $N$ , we can merge the equations of motion in the following system:

$$\dot{z} = \mathbf{h} + G\Gamma \tag{1.63}$$

where

$$\dot{z} = \begin{pmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \vdots \\ \dot{z}_N \end{pmatrix}, \quad \mathbf{h} = \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_N \end{pmatrix}, \quad G = \begin{pmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,N} \\ g_{2,1} & g_{2,2} & \cdots & g_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ g_{N,1} & g_{N,2} & \cdots & g_{N,N} \end{pmatrix}, \quad \Gamma = \begin{pmatrix} \Gamma_1 \\ \Gamma_2 \\ \vdots \\ \Gamma_N \end{pmatrix}$$

Therefore, we will have  $N$  different  $D^{(1)}$ 's and  $N^2$  different  $D^{(2)}$ 's.

### 1.3.2 Exercise

Consider now the case of a Brownian particle in a potential

$$\begin{cases} \ddot{x} + \gamma\dot{x} + f'(x) = \sigma\Gamma(t) \\ \mathbb{E}[\Gamma(t)] = 0 \\ \mathbb{E}[\Gamma(t)\Gamma(t')] = \delta(t - t') \end{cases} \tag{1.64}$$

To use the Fokker-Planck equation, we need first order equations. To remove the second order term  $\ddot{x}$  let us use a system of two equations, defining  $\dot{x}$  as  $v$

$$\begin{cases} \dot{x}(t) = v(t) \\ \dot{v}(t) = -\gamma v(t) - f'(x(t)) + \sigma\Gamma(t) \end{cases} \tag{1.65}$$

Therefore,

$$\begin{aligned} h_x &= v(t) & g_{xx} &= 0 & g_{xv} &= 0 \\ h_v &= -\gamma v(t) - f'(x(t)) & g_{vx} &= 0 & g_{vv} &= \sigma \end{aligned}$$

and

$$\begin{aligned} D_x^{(1)} &= v(t) & D_v^{(1)} &= -\gamma v(t) - f'(x(t)) \\ D_{xx}^{(2)} &= 0 & D_{xv}^{(2)} &= 0 & D_{vx}^{(2)} &= 0 & D_{vv}^{(2)} &= \sigma^2 \end{aligned}$$

For Fokker-Planck

$$\begin{aligned} \frac{\partial W}{\partial t} &= \left[ -\frac{\partial}{\partial x} D_x^{(1)} - \frac{\partial}{\partial v} D_v^{(1)} + \frac{\partial^2}{\partial x^2} D_{xx}^{(2)} + \frac{\partial^2}{\partial x \partial v} D_{xv}^{(2)} + \frac{\partial^2}{\partial v \partial x} D_{vx}^{(2)} + \frac{\partial^2}{\partial v^2} D_{vv}^{(2)} \right] W \\ &= -\frac{\partial}{\partial x} v W - \frac{\partial}{\partial v} (-\gamma v(t) - f'(x(t))) W + \frac{\partial^2}{\partial v^2} \sigma^2 W \end{aligned} \quad (1.66)$$

If  $\sigma^2 = \gamma \frac{k_B T}{m} = \gamma v_{th}^2$  and  $f'(x) = w_o^2 x$ ,

$$\frac{\partial W}{\partial t} = -v \frac{\partial W}{\partial x} + \gamma \left( W + v \frac{\partial W}{\partial v} \right) + w_o^2 x \frac{\partial W}{\partial v} + \gamma v_{th}^2 \frac{\partial^2 W}{\partial v^2} \quad (1.67)$$

the solution for the stationary case is

$$W_{st}(x, v) = \frac{w_o}{2\pi v_{th}^2} e^{-\frac{1}{2} \frac{v^2}{v_{th}^2}} e^{-\frac{1}{2} \frac{w_o^2 x^2}{v_{th}^2}} = \frac{m w_o}{2\pi k_B T} e^{-\frac{mv^2 + mw_o^2 x^2}{2k_B T}} = \frac{m w_o}{2\pi k_B T} e^{-\frac{E}{k_B T}} \quad (1.68)$$

By integration, we obtain the stationary densities for the various observables:

$$W_{st}(x) = \frac{m w_o}{2\pi k_B T} e^{-\frac{mw_o^2 x^2}{2k_B T}} \int_{-\infty}^{+\infty} e^{-\frac{mv^2}{2k_B T}} dv \quad (1.69)$$

Substituting

$$\frac{mv^2}{k_B T} = \frac{u^2}{s^2} \Rightarrow du = s \sqrt{\frac{m}{k_B T}} dv \Rightarrow dv = \sqrt{\frac{k_B T}{ms^2}} du \quad (1.70)$$

Therefore,

$$W_{st}(x) = \frac{m w_o}{2\pi k_B T} e^{-\frac{mw_o^2 x^2}{2k_B T}} \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2s^2}} \sqrt{\frac{k_B T}{ms^2}} du = \frac{m w_o}{2\pi k_B T} e^{-\frac{mw_o^2 x^2}{2k_B T}} \sqrt{\frac{2\pi k_B T}{m}} = \sqrt{\frac{m w_o^2}{2\pi k_B T}} e^{-\frac{mw_o^2 x^2}{2k_B T}} \quad (1.71)$$

Similarly

$$W_{st}(v) = \frac{m w_o}{2\pi k_B T} e^{-\frac{mv^2}{2k_B T}} \int_{-\infty}^{+\infty} e^{-\frac{mw_o^2 x^2}{2k_B T}} dx = \sqrt{\frac{m}{2\pi k_B T}} e^{-\frac{mv^2}{2k_B T}} \quad (1.72)$$

This implies that

$$\begin{aligned}\mathbb{E}[x] &= 0; & \mathbb{E}[v] &= 0 \\ \mathbb{E}[x^2] &= \frac{k_B T}{m w_o^2}; & \mathbb{E}[v^2] &= \frac{k_B T}{m}\end{aligned}$$

### 1.3.3 Exercise

Consider the Langevin representation of the gravitational wave detector introduced in Ref. [11]:

$$\begin{cases} \dot{q}(t) = I(t) \\ \dot{I}(t) = -\frac{R_k + R_d}{L_k} I - \frac{1}{L_k C_k} q + \sqrt{2k_B T_{eff}(R_k + R_d)} \Gamma \end{cases} \quad (1.73)$$

which implies

$$\gamma = \frac{R_k + R_d}{L_k}; \quad w_o^2 = \frac{1}{L_k C_k}; \quad \sigma = \sqrt{2k_B T_{eff}(R_d + R_k)}$$

As we took  $\sigma^2 = \frac{\gamma k_B T}{m}$ , the above averages can be written as

$$\begin{aligned}\mathbb{E}[x] &= 0; & \mathbb{E}[v] &= 0 \\ \mathbb{E}[x^2] &= \frac{\sigma^2}{\gamma w_o^2}; & \mathbb{E}[v^2] &= \frac{\sigma^2}{\gamma}\end{aligned}$$

Consequently, our specific model has Gaussian distributions with

$$\mathbb{E}[q] = 0; \quad \mathbb{E}[I] = 0 \quad (1.74)$$

and

$$\begin{aligned}\mathbb{E}[q^2] &= \text{var}(q) = \frac{\sigma^2}{\gamma w_o^2} = \frac{2k_B T_{eff}(R_k + R_d)}{\frac{(R_k + R_d)}{L_k} \frac{1}{L_k C_k}} = 2k_B T_{eff} L_k^2 C_k \\ \mathbb{E}[I^2] &= \text{var}(I) = \frac{\sigma^2}{\gamma} = \frac{2k_B T_{eff}(R_k + R_d)}{\frac{(R_k + R_d)}{L_k}} = 2k_B T_{eff} L_k\end{aligned}$$

To obtain the PDF's of

$$q_\tau(t) = \frac{1}{\tau} \int_t^{t+\tau} q(s) ds \quad \text{and of} \quad I_\tau(t) = \frac{1}{\tau} \int_t^{t+\tau} I(s) ds \quad (1.75)$$

consider that we have a Gaussian process, hence the PDF's of its variables observed at different times are jointly Gaussian.

## 1.4 Fluctuation-Dissipation Theorem

In general, let  $Z(t)$  be a stationary process, which is described for a time  $t \in [0, T]$ . We may pretend  $Z$  is periodic of period  $T$ , because  $T$  can be taken arbitrarily large; larger, in particular, than the physically interesting time scales. Under such assumption, one can expand the values  $z$  that the variable  $Z$  takes, as:

$$z(t) = \sum_{n=-\infty}^{+\infty} a_n e^{i\omega_n t}; \quad \text{where } a_n = \frac{1}{T} \int_0^T z(t) e^{-i\omega_n t} dt; \quad \omega_n = \frac{2\pi n}{T} \quad (1.76)$$

If  $T$  is large, and  $z$  sufficiently regular,  $\omega_n$  can be considered almost continuous in  $n$ , for the relevant  $n$ . Given  $a_n = \alpha_n + i\beta_n$  with  $\alpha, \beta \in \mathbb{R}$ , one needs  $a_{-n} = a_n^* = \alpha_n - i\beta_n$  for  $Z$  to be real. As  $Z$  is a random variable, so are its Fourier coefficients  $a_n$ ; their averages over the realizations of the process can be expressed as:

$$\begin{aligned} \mathbb{E}[a_0] &= \frac{1}{T} \int_0^T \mathbb{E}[z(t)] dt = \mathbb{E}[Z] \\ \mathbb{E}[a_n] &= \frac{1}{T} \int_0^T \mathbb{E}[z(t)] e^{-i\omega_n t} dt = 0 \quad \text{if } n \neq 0 \end{aligned}$$

where we have used that the process is stationary, hence, by definition,  $\mathbb{E}[z(t)] = \mathbb{E}[Z]$ . The time average for a single realization of the process, i.e. for one given set of coefficients  $\{a_n\}_{-\infty}^{\infty}$ , yields:

$$\bar{z}^T = \frac{1}{T} \int_0^T z(t) dt = a_0 \quad (1.77)$$

which in general differs from  $\mathbb{E}[a_0]$ . The process is called *ergodic* if:

$$\lim_{T \rightarrow \infty} \bar{z}^T = \mathbb{E}[Z] \quad (1.78)$$

Commonly, stochastic processes are also ergodic, hence this is assumed to be the case, so that  $a_0 = \mathbb{E}[a_0] = \mathbb{E}[Z]$ . Then, the Fourier coefficients  $b_n$  of the process  $Y(t) = Z(t) - \mathbb{E}[Z]$  obey:

$$\mathbb{E}[b_n] = 0 \quad \forall n \in \mathbb{Z} \quad (1.79)$$



which can be assumed to be the case in general, since the subtraction of a constant from a stochastic process does not alter its significance.

The average strength of the Fourier component  $a_n$  is defined by:

$$\mathbb{E}[|a_n|^2] = \mathbb{E}[\alpha_n^2] + \mathbb{E}[\beta_n^2] \quad (1.80)$$

and the average intensity  $I_T$  around frequency  $\omega$  is defined by:

$$I_T(\omega; \Delta\omega)\Delta\omega = \sum_{\omega_n \in [\omega, \omega + \Delta\omega]} \mathbb{E}[|a_n|^2] \quad (1.81)$$

where  $\Delta\omega$  must be larger than the difference of two consecutive angular frequencies,  $2\pi/T$  for the interval  $[\omega, \omega + \Delta\omega)$  to contain at least one angular frequency. In the case in which  $T$  is very large,  $\Delta\omega$  can be small, and the Fourier components indexed by the angular frequencies within  $[\omega, \omega + \Delta\omega)$  are approximately equal, if the process is sufficiently regular. One may then write:

$$I_T(\omega; \Delta\omega)\Delta\omega \approx \mathbb{E}[|a_\omega|^2] \frac{T}{2\pi} \Delta\omega \quad (1.82)$$

where  $\hat{\omega}$  is any angular frequency in  $[\omega, \omega + \Delta\omega)$ , e.g. the smallest, and  $T/2\pi$  is the number of frequencies per frequency unit, so that  $T\Delta\omega/2\pi$  for very large  $T$  approximately equals the number of frequencies in  $[\omega, \omega + \Delta\omega)$ .

Given the process  $Z$  and the accepted tolerance  $\Delta\omega$ , one may choose  $T$  large enough that the approximate equalities can be treated as equalities, and  $a_n$  as a practically continuous function of  $\omega$ . Mathematically, this requires the large  $T$  limit with fixed  $zD\omega$  to be taken first, so that arbitrarily small  $\Delta\omega$  can be taken after, for  $a_n$  to approximate better and better a given value  $a_\omega$ . As  $T$  increases, the power of the signal in a the fixed interval  $[\omega, \omega + \Delta\omega)$  should reach a finite value, therefore the power of each single frequency must tend to 0. If this is the case, and it is the case for standard physical applications, one may divide by  $\Delta\omega$  and take the  $\Delta\omega \rightarrow 0$  limit, thus obtaining the intensity spectrum of  $Z$  as:

$$I(\omega) = \lim_{\substack{\Delta\omega \rightarrow 0 \\ \hat{\omega} \in [\omega, \omega + \Delta\omega)}} \lim_{T \rightarrow \infty} I_T(\hat{\omega}, \Delta\omega) = \lim_{\substack{\Delta\omega \rightarrow 0 \\ \hat{\omega} \in [\omega, \omega + \Delta\omega)}} \lim_{T \rightarrow \infty} \frac{T}{2\pi} \mathbb{E}[|a_{\hat{\omega}}|^2] \quad (1.83)$$

The terminology can be understood thinking about time dependent currents in electrical circuits. Letting  $I$  be the current,  $R$  the resistance and  $V$  the electric potential, the dissipated power  $P$  under Ohm's law is given by

$$P(t) = I(t)V(t) = R|I(t)|^2 \quad (1.84)$$

Expressing  $I$  in terms of its Fourier components,  $I(t) = \sum_n a_n \exp(i\omega_n t)$ , the average dissipated power then becomes:

$$\bar{P} = \frac{1}{T} \int_0^T I(t)I^*(t)dt = \frac{1}{T} \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} a_n a_m^* \int_0^T e^{i(\omega_n - \omega_m)t} dt = \sum_{n=-\infty}^{\infty} |a_n|^2 \quad (1.85)$$

This shows that the square modulus of a Fourier component of  $I$  corresponds to the average power dissipated at angular frequency  $\omega_n$ .

Introduce the autocorrelation  $\Phi_Z(t_1, t_2) = \mathbb{E}[Z(t_1)Z(t_2)]$  of  $Z$ , which can be written as  $\Phi_Z(t_1, t_2) = \mathbb{E}[Z(t_1)Z(t_1 + t)]$ , where  $t = t_2 - t_1$ , to stress that it depends on the initial time, and on the difference between initial and final observation time. In a stationary state, the initial time makes no difference, therefore one may simply write  $\Phi_Z(t) = \mathbb{E}[Z(t_0)Z(t_0 + t)]$ , where  $t_0$  is any initial time. Then, the *Wiener-Khinchin theorem* asserts that the spectral power density of  $Z$ ,  $I_Z$ , and its autocorrelation  $\Phi_Z$  are each other's Fourier transforms:

$$I_Z(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \Phi_Z(t) e^{-i\omega t} dt, \quad \Phi_Z(t) = \int_{-\infty}^{+\infty} I_Z(\omega) e^{i\omega t} d\omega \quad (1.86)$$

i.e. knowledge of the power spectrum is equivalent to knowledge of the autocorrelation.

These notions can be applied to the Brownian motion, that obeys:

$$m\dot{v} = -m\gamma v + m\Gamma \quad \text{or} \quad \dot{v} = -\gamma v + \Gamma \quad (1.87)$$

if the velocity and the stochastic term are expanded as:

$$\Gamma(t) = \sum_{n=-\infty}^{+\infty} \Gamma_n e^{i\omega_n t}, \quad v(t) = \sum_{n=-\infty}^{+\infty} v_n e^{i\omega_n t} \quad (1.88)$$

where  $v_n$  is the Fourier coefficient of  $v$  and  $\Gamma_n$  that of  $\Gamma$ , at frequency  $\omega_n$ . Then, substituting (1.88) in (1.87), one obtains:

$$\sum_{n=-\infty}^{+\infty} i\omega_n v_n e^{i\omega_n t} = - \sum_{n=-\infty}^{+\infty} \gamma v_n e^{i\omega_n t} + \sum_{n=-\infty}^{+\infty} \Gamma_n e^{i\omega_n t} \quad (1.89)$$

which holds for all times  $t \in [0, T]$  if and only if

$$i\omega_n v_n = -\gamma v_n + \Gamma_n \quad \text{i.e.} \quad v_n = \frac{\Gamma_n}{i\omega_n + \gamma} \quad (1.90)$$

Then, for the power spectra of  $v$  and  $\Gamma$ , we get:

$$I_v(\omega) = \lim_{\substack{\Delta\omega \rightarrow 0 \\ \hat{\omega} \in [\omega, \omega + \Delta\omega]}} \lim_{T \rightarrow \infty} \frac{T}{2\pi} \mathbb{E} \left[ \frac{|\Gamma_{\hat{\omega}}|^2}{\hat{\omega}^2 + \gamma^2} \right] = \frac{I_\Gamma(\omega)}{\omega^2 + \gamma^2} \quad (1.91)$$

In the simplest case,  $I_\Gamma(\omega) = I_\Gamma = \text{constant}$ , in which all Fourier components contribute the same power, the spectrum of  $\Gamma$  is called white, and the spectrum of  $v$ ,  $I_v$ , is Lorentzian:

$$I_v(\omega) = \frac{I_\Gamma}{\omega^2 + \gamma^2} \quad (1.92)$$

In a stationary state, taking  $t_1 > t_2$ , one gets:

$$\Phi_\Gamma(t_1 - t_2) = \mathbb{E}[\Gamma(t_1)\Gamma(t_2)] = \int_{-\infty}^{+\infty} I_\Gamma e^{i\omega(t_1-t_2)} d\omega \quad (1.93)$$

Considering a stationary state, taking  $t_0 = t_2$ , and using the following representation of the Dirac delta function:

$$\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ikx} dk, \quad \text{which means} \quad \int_{-\infty}^{+\infty} dx \frac{1}{2\pi} \int_{-\infty}^{+\infty} dk f(x) e^{ikx} = f(0) \quad (1.94)$$

one gets:

$$\Phi_\Gamma(t_1 - t_2) = 2\pi I_\Gamma \delta(t_1 - t_2) \quad (1.95)$$

which shows that  $q = 2\pi I_\Gamma$  in the Langevin treatment of Sect. 1.3.1. Analogously, the autocorrelation of  $v$  reads:

$$\Phi_v(t) = \int_{-\infty}^{+\infty} \frac{I_\Gamma}{\omega^2 + \gamma^2} e^{i\omega t} d\omega \quad (1.96)$$

Then, for  $t = t_1 - t_2$ , one may write:

$$\Phi_v(t_1 - t_2) = \mathbb{E}[v(t_1)v(t_2)] = I_\Gamma \int_{-\infty}^{+\infty} \frac{e^{i\omega(t_1-t_2)}}{\omega^2 + \gamma^2} d\omega = I_\Gamma \frac{\pi}{\gamma} e^{-\gamma(t_1-t_2)} \quad \text{for } t_1 > t_2 \quad (1.97)$$

or

$$\mathbb{E}[v(t_1)v(t_2)] = \frac{\pi I_\Gamma}{\gamma} e^{-\gamma|t_1-t_2|}, \quad \forall t_1, t_2 \in \mathbb{R} \quad (1.98)$$

So velocity correlations decay exponentially in time. Then, taking  $t_1 = t_2 = t$ , and  $t$  large enough that the steady state has been reached, and averages do not depend on time anymore, one obtains:

$$\mathbb{E}[v^2] = \frac{\pi I_\Gamma}{\gamma} \Rightarrow I_\Gamma = \frac{\gamma}{\pi} \mathbb{E}[v^2] \quad (1.99)$$

If the Brownian particle has equilibrated with the fluid, equipartition of energy applies,  $m\mathbb{E}[v^2] = k_B T$ , hence

$$I_\Gamma = \frac{\gamma k_B T}{\pi m} \tag{1.100}$$

in accord with Eq. (1.31).

Equation (1.100) states that the power of the random force is proportional to the friction coefficient and to the thermal energy. In turn, Einstein's relation (1.37) represents an inverse proportionality between the diffusion coefficient and the friction coefficient. Both results demonstrate that the dissipation of energy caused by friction is intimately related to the equilibrium fluctuations, due to the incessant molecular motions.

### 1.4.1 Exercise: Derivation of Eq. (1.98)

Consider the inverse  $\mathcal{F}$ -transform of a given function  $f$ :

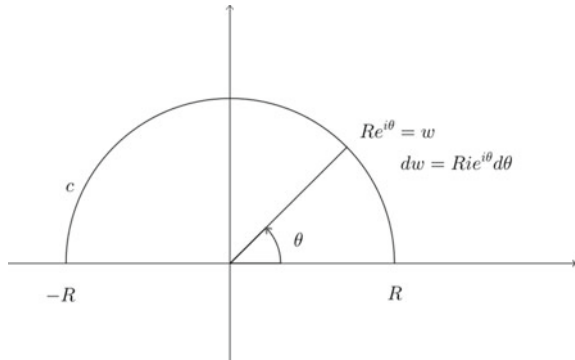
$$\int_{-\infty}^{+\infty} f(\omega) e^{i\omega t} d\omega \tag{1.101}$$

where  $t \in \mathbb{R}_+$  and  $f$  is analytic in the upper half plane, except for a finite number of poles, and assume  $\lim_{|\omega| \rightarrow \infty} f(\omega) = 0$ , with  $\arg \omega \in [0, \pi]$  (Fig. 1.5). Recall that

$$\oint_c F(\omega) d\omega = \int_{-R}^R F(x) dx + \int_0^\pi F(Re^{i\theta}) i Re^{i\theta} d\theta = 2\pi i \sum \text{residues in upper half plane} \tag{1.102}$$

and Cauchy's formula

**Fig. 1.5** Circuit for integration in the complex plane



$$\oint_c \frac{F(\omega)}{\omega - \omega_0} d\omega = 2\pi i F(\omega_0) \tag{1.103}$$

if  $\omega_0$  within  $c$ . Introduce the finite  $R$  approximation of the Fourier anti-transform of  $f$ , then consider

$$I_R = \int_0^\pi f(Re^{i\theta}) e^{it(R \cos \theta + iR \sin \theta)} i R e^{i\theta} d\theta \tag{1.104}$$

Let  $R$  be large enough that  $|f(Re^{i\theta})| < \varepsilon$ . Therefore

$$|I_R| \leq \varepsilon R \int_0^\pi \left| e^{itR \cos \theta} e^{-tR \sin \theta} e^{i\frac{\pi}{2}} e^{i\theta} \right| d\theta = \varepsilon R \int_0^\pi e^{-tR \sin \theta} d\theta = 2\varepsilon R \int_0^{\pi/2} e^{-tR \sin \theta} d\theta \tag{1.105}$$

Now,  $\frac{2}{\pi}\theta \leq \sin \theta$  for  $\theta \in [0, \pi/2]$ , hence

$$|I_R| \leq 2\varepsilon R \int_0^{\pi/2} e^{-tR \frac{2}{\pi}\theta} d\theta = 2\varepsilon R \left. \frac{e^{-tR2\theta/\pi}}{-tR2/\pi} \right|_0^{\pi/2} = \varepsilon \frac{1 - e^{-tR}}{t\pi} = \frac{\pi\varepsilon}{t} (1 - e^{-tR}) \tag{1.106}$$

Because  $\varepsilon$  is arbitrarily small (it suffices to take  $R$  large), we have

$$\lim_{R \rightarrow \infty} \int_0^\pi f(Re^{i\theta}) e^{itR(\cos \theta + i \sin \theta)} R e^{i\theta} i d\theta = 0 \tag{1.107}$$

Hence

$$\int_{-\infty}^{+\infty} f(\omega) e^{i\omega t} d\omega = \lim_{R \rightarrow \infty} \left[ \int_{-R}^R f(\omega) e^{i\omega t} d\omega + \int_0^\pi f(Re^{i\theta}) e^{itRe^{i\theta}} i R e^{i\theta} d\theta \right] \tag{1.108}$$

$= 2\pi i \sum$  upper half plane residues

if  $t > 0$  ( $c$  oriented counterclockwise).

Now, the residue of a pole of order  $m$  of a given  $F$  is defined by

$$a_{-m} = \frac{1}{(m-1)!} \left. \frac{d^{m-1}}{dz^{m-1}} [(z - z_0)^m F(z)] \right|_{z=z_0} \tag{1.109}$$

For example

$$a_{-1} = (z - z_0) F(z) \Big|_{z_0} \tag{1.110}$$

Clearly,  $\frac{e^{i\omega t}}{\omega^2 + \gamma^2} = \frac{e^{i\omega t}}{(\omega + i\gamma)(\omega - i\gamma)}$  has only one pole in the upper half plane:  $w_0 = i\gamma$  and  $f(\omega) \xrightarrow{|\omega| \rightarrow \infty} 0$ . Therefore, we can apply the above calculations:

$$a_{-1} = (\omega - i\gamma) \frac{e^{i\omega t}}{\omega^2 + \gamma^2} \Big|_{i\gamma} = \frac{e^{i\omega t}}{\omega + i\gamma} \Big|_{i\gamma} = \frac{e^{-\gamma t}}{2i\gamma} \tag{1.111}$$

Hence

$$\int_{-\infty}^{+\infty} \frac{e^{i\omega t}}{\omega^2 + \gamma^2} d\omega = \pi \frac{e^{-\gamma t}}{\gamma} = \frac{\pi}{\gamma} e^{-\gamma t} \tag{1.112}$$

For negative  $t$ , use the clockwise contour and obtain  $\frac{\pi}{\gamma} e^{\gamma t} = \frac{\pi}{\gamma} e^{-\gamma|t|}$ . The negative sign that emerges in the denominator is cancelled by the clockwise rotation.

### 1.4.2 Application: Johnson-Nyquist Noise

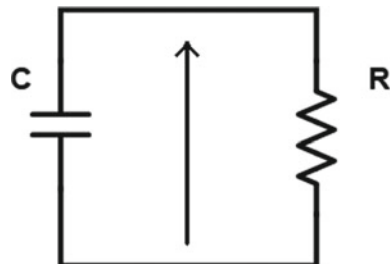
Consider an  $RC$ -circuit in equilibrium at temperature  $T$ , with a time dependent potential difference  $U(t)$  at the ends of its resistance. This describes an experiment performed by Johnson, explained by Nyquist in 1928. It is indeed possible to observe 1 electron at a time is emitted by a hot filament (Schottky). Because of their temperature, the electrons in the circuit move to up and down in it, but with equal probabilities, so there is no net current, on average (Fig. 1.6).

However, on short time scales, one may find more electrons going up than down, and vice-versa. Thus  $U(t)$  fluctuates with zero mean. The variance of these fluctuations can only be related to the value of  $T$ . Let  $Q(t) = CU(t)$  be the charge in the capacitor, and try this description:

$$R \frac{dQ}{dt} = -\frac{1}{C} Q + \eta(t); \quad I(t) = \frac{dQ}{dt}(t) \Rightarrow \dot{Q} + \frac{1}{RC} Q = \frac{1}{R} \eta; \quad \dot{U} + \frac{1}{RC^2} U = \frac{1}{RC} \eta \tag{1.113}$$

If we take  $\mathbb{E}[\eta(t)\eta(t')] = 2Rk_B T \delta(t - t')$  we have a Langevin equation formally identical to the one for the Brownian motion  $\dot{v} + \gamma v = \Gamma$ , with  $\mathbb{E}[\Gamma(t)\Gamma(t')] = q\delta(t - t')$ . In particular, comparing the two, one may write:

**Fig. 1.6**  $RC$ -circuit with time dependent current due to a time dependent potential  $U$



$$\begin{aligned}\mathbb{E}\left[\frac{\eta(t)}{R}\frac{\eta(t')}{R}\right] &= \frac{2k_B T}{R}\delta(t-t') \Rightarrow q_Q = \frac{2k_B T}{R} \\ \mathbb{E}\left[\frac{\eta(t)}{RC}\frac{\eta(t')}{RC}\right] &= \frac{2k_B T}{RC^2}\delta(t-t') \Rightarrow q_U = \frac{2k_B T}{RC^2}\end{aligned}\quad (1.114)$$

Then, as

$$\begin{aligned}\mathbb{E}[v] = 0 &\Rightarrow \mathbb{E}[Q] = 0 \Rightarrow \mathbb{E}[U] = 0 \\ \mathbb{E}[v^2] = \frac{q}{2\gamma} = \frac{k_B T}{m} &\Rightarrow \mathbb{E}[Q^2] = \frac{q_Q}{2\gamma} = Ck_B T \Rightarrow \mathbb{E}[U^2] = \frac{q_U}{2\gamma} = \frac{k_B T}{C}\end{aligned}\quad (1.115)$$

We also have  $\mathbb{E}[Q(t)Q(t')] = k_B T C e^{-|t-t'|/RC}$  etc. and from

$$I_v(\omega) = \frac{2\gamma k_B T}{m(\omega^2 + \gamma^2)} = \frac{q}{\gamma^2(1 + \frac{\omega^2}{\gamma^2})}\quad (1.116)$$

we obtain

$$I_U(\omega) = \frac{\frac{q_U}{\gamma^2}}{1 + \frac{\omega^2}{\gamma^2}} = \frac{2Rk_B T}{1 + (RC\omega)^2}\quad (1.117)$$

which, for low frequencies  $\omega \ll (RC)^{-1}$ , implies that  $I_U$  does not depend on  $C$ :  $I_U(\omega) \approx 2Rk_B T$ . This has been experimentally confirmed, demonstrating the incredible success of the Brownian motion theory in the most diverse phenomena.

### 1.4.3 Recent Variations of Brownian Motion: Active Matter

Unlike molecules, or inert objects in general, *active particles* propel themselves thanks to various kinds of mechanisms: thanks to *flagella*, for instance, certain kinds of bacteria enjoy self mobility. If such active particles are immersed in a fluid, one possible representation of their dynamics is a variation of the BM, in which particles are endowed with a kind of engine. Indeed, dimensions of bacteria are even smaller than that of pollen, hence in principle they should be similarly affected by molecular impacts. In Ref. [12], the following model of a 2-dimensional active fluid is thus proposed:

$$\dot{x} = -\mu_t \partial_x U(x) + \nu \cos \vartheta + \sqrt{2D_t} \Gamma_x \quad (1.118)$$

$$\dot{y} = \nu \sin \vartheta + \sqrt{2D_t} \Gamma_y \quad (1.119)$$

$$\dot{\vartheta} = \mu_r G(\vartheta) + \sqrt{2D_r} \Gamma_\theta, \quad (1.120)$$

where  $(x, y)$  is the position (e.g. enter of mass) of an active particle,  $\vartheta$  is its orientation with respect to  $x$ -axis,  $U(x)$  is a potential eprresenting a wall,  $G$  is the torque this wall produces,  $D_t, D_r$  are diffusivities,  $\mu_t, \mu_r$  are two coefficients related to friction,  $\Gamma_x, \Gamma_y, \Gamma_\theta$  are independent Gaussian noises, and  $\nu$  is the activity coefficient, which models the effect of the engine of the particle. In this model, the engine pushes the particle in the directions of its orientation. The paper consideres elliptical particles (with semi-axes  $a$  and  $b$ ) and a harmonic potential,

$$U = \frac{\lambda}{2} \Theta[x - x_w](x - x_w)^2, \quad (1.121)$$

where  $\Theta$  is the Heaviside function and  $x_w$  is the wall position. From (1.121) it follows that the torque reads

$$G = m\lambda\kappa\Theta[x - x_w] \sin 2\vartheta, \quad (1.122)$$

with  $\kappa = (a^2 - b^2)/8$ . This overdamped model is associated with the following Fokker-Plank equation:

$$\partial_t \mathcal{P} = -\nabla[(\mathbf{v} - \mu_t \nabla U(x))\mathcal{P} - D_t \nabla \mathcal{P}] - \partial_\theta[\mu_r G(x, \theta)\mathcal{P} - D_r \partial_\theta \mathcal{P}] \quad (1.123)$$

In Ref. [12], the pressure is then defined as:

$$P = -\left. \frac{\partial F}{\partial L} \right|_N \quad (1.124)$$

where  $L$  is the system length, the number of active particles  $N$  is kept constant, and  $F$  is the free energy defined by:

$$F = -\frac{1}{\beta} \ln \mathcal{Z} \quad (1.125)$$

where

$$\mathcal{Z} = \sum_n e^{-\beta[\mathcal{H} + \sum_i U(x_i - L)]} \quad (1.126)$$

is the partition function,  $\beta = \frac{1}{T}$ ,  $T$  being the temperature, and the sum runs over all micro-states. The origin of the wall is set at  $x = L = x_w$ ,  $U(x_i - L)$  is the wall potential with  $x_i$  the position of particle  $i$ , and  $\mathcal{H}$  contains all the other interactions in the sysyem. Then, one can write:

$$P = -\frac{1}{\mathcal{Z}} \sum_n \sum_i \partial_L U(x_i - L) e^{-\beta[\mathcal{H} + \sum_i U(x_i - L)]} = -\left\langle \int dx \rho(x) \partial_L U(x - L) \right\rangle \quad (1.127)$$



where the angular brackets denote a thermal average, and  $\rho(x) = \sum_i \delta(x - x_i)$  is the number density. Exchanging  $\partial_L$  for  $-\partial_x$ , one obtains:

$$P = \left\langle \int dx \rho(x) \partial_x U(x - L) \right\rangle \quad (1.128)$$

The authors of Ref. [12] then conclude that this system does not have an intrinsic pressure, because it depends on the wall potential confining the system.

There are various open questions concerning this result. In the first place, the applicability of the overdamped picture for particles equipped with an internal engine should be investigated; moreover the role of the canonical expression for the pressure should also be analyzed. In fact, unlike the BM, activity of particles leads to a net dissipation, i.e. to a nonequilibrium state. One may thus consider an underdamped picture, like the following:

$$\dot{x} = v_x, \quad (1.129)$$

$$\dot{y} = v_y, \quad (1.130)$$

$$\dot{\theta} = v_\theta, \quad (1.131)$$

$$\dot{v}_x = -\frac{1}{m} \frac{\partial U}{\partial x} + \frac{\nu \cos \frac{\theta}{\sqrt{J}}}{m\mu_t} - \frac{v_x}{m\mu_t} + \frac{\sqrt{2D_t}}{m\mu_t} \Gamma_x, \quad (1.132)$$

$$\dot{v}_y = -\frac{1}{m} \frac{\partial U}{\partial y} + \frac{\nu \sin \frac{\theta}{\sqrt{J}}}{m\mu_t} - \frac{v_y}{m\mu_t} + \frac{\sqrt{2D_t}}{m\mu_t} \Gamma_y, \quad (1.133)$$

$$\dot{v}_\theta = \frac{G[\frac{\theta}{\sqrt{J}}]}{mJ^{1/2}} - \frac{v_\theta}{\mu_r m J} + \frac{\sqrt{2D_r}}{\mu_r m J^{1/2}} \Gamma_\theta, \quad (1.134)$$

where we have introduced the inertial moment per unit mass of the elliptical shape,  $J = (a^2 + b^2)/5$  and we have rescaled the orientation angle as  $\theta = \vartheta J^{1/2}$  (and accordingly  $\Gamma_\theta = \Gamma_\vartheta$ ). In principle, the underdamped picture should converge to the overdamped picture, when inertia, i.e. the mass of particles become small. However, this limit is known to be singular, in general, and the activity of particles may make it even more troublesome. This second approach then gives access to different equations for the probability distribution of active particles, and to different expressions for the thermodynamic quantities, pressure included [13]. We may safely argue that the investigation of such and many other tremendously interesting questions has just begun [14].

### 1.4.4 Recent Variations of Brownian Motion: Detection of Gravitational Waves

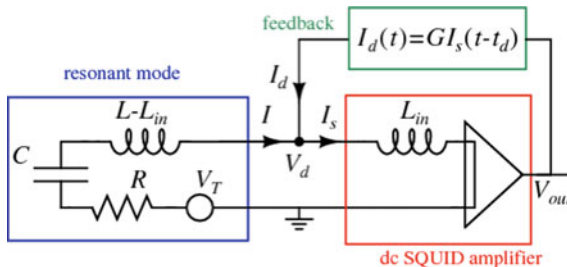
One second variation of the BM gives us a glimpse of the breadth of phenomena to which this kind of models has been successfully applied: the study of noise in gravitational waves detectors [11]. Here we consider the resonant bars, which are one kind of detectors that are supposed to resonate with gravitational waves, when they pass. In the absence of gravitational waves, these bars vibrate and, like any solid in an equilibrium state at a given temperature  $T$ , the variance of such vibrations is proportional to  $T$ . This is the thermal noise which a classical macroscopic object cannot avoid, and interferes with the vibrations induced by the gravitational waves. However, it can be reduced by reducing  $T$ . In particular, the bar known as AURIGA is an aluminum bar of mass  $2.2 \times 10^3$  kg and length of 3 m, which is cooled to liquid helium temperature  $T_0 = (4.6 \pm 0.2)$  K. Then, in order to improve the stability of the device, its modes of vibration are further “cooled” by a feedback mechanism, which acts as a kind of viscosity.

Such an experimental device can be described through separate oscillators normal modes, each of which behaves like an RLC series electrical circuit. Denoting the effective different inductance of a mode by  $L$ , the capacitance by  $C$  and the resistance by  $R$ , cf. Fig. 1.7, the corresponding evolution equations take the form:

$$(L - L_{in}) \frac{d^2 q(t)}{dt^2} + R \frac{dq(t)}{dt} + \frac{q(t)}{C} = V_T(t) - V_d(t) \quad (1.135a)$$

$$V_d(t) = L_{in} \frac{dI_s(t)}{dt}, \quad \text{where } I(t) + I_d(t) = I_s(t) \quad (1.135b)$$

Here,  $q$  is the charge on the capacitor,  $I = dq(t)/dt$  is the current through the inductance  $L$ ,  $V_d$  is the voltage at the node where the feedback takes place, and  $L_{in}$  is the input inductance of the SQUID amplifier. The feedback form chosen in Ref. [11] is:



**Fig. 1.7** RLC circuit describing one mode of vibration of a feedback cooled resonant bar. A dc SQUID is represented as current amplifier. The observable is the current  $I_s$ , and the feedback corresponds to a current  $I_d$  which is a delayed copy of  $I_s$  reduced by a factor  $G \ll 1$ . The SQUID output voltage is  $V_{out} = A I_s$

$$I_d(t) = GI_s(t - t_d), \quad \text{with } t_d = \frac{\pi}{2\omega_r} \text{ and } G \ll 1 \quad (1.136)$$

It is now assumed that each oscillator is driven by a stochastic voltage of the form  $V_T(t) = \sqrt{2k_B T_0 R} \Gamma(t)$ , where  $\Gamma$  is a Gaussian noise. Moreover, the very high quality factor of the oscillators imply that the currents  $I_s$ ,  $I_d$  and  $I$  have frequency quite close to  $\omega_r = 1/\sqrt{LC}$ , while their amplitudes and phases change very slowly. Therefore, the following approximation

$$I(t) = \hat{I}(t) \sin[\omega_r t + \hat{\phi}(t)]$$

and the analogous ones for the other currents result quite accurate. Now, although Eq. (1.136) implies memory effects, due to contributions from times  $t - t_d$ , the above quasi-harmonic approximation implies  $I_s(t - t_d) \simeq \omega_r q_s(t)$ , and  $I_s(t) = dq_s(t)/dt$  obeys:

$$L \frac{dI_s(t)}{dt} + I_s(t) [R + R_d] + \frac{q_s(t)}{C} = \sqrt{2k_B T_0 R} \Gamma(t) \quad (1.137)$$

with  $R_d = G\omega_r L_{in}$  playing the role of the viscous damping. Introducing  $g = R_d/R$ , as the efficiency of the feedback mechanism, Eq.(1.137) turns identical to the Langevin equation with damping  $R + R_d$ , in an equilibrium state at the fictitious temperature  $T_{\text{eff}} = T_0/(1 + g)$ .

While the fact that  $T_{\text{eff}}$  differs from the bath temperature  $T_0$  reveals the nonequilibrium nature of the phenomenon, one may formally solve Eq. (1.137) as usual in the equilibrium case. This has been done in Ref. [11], and the analytical as well as the numerical simulations results have been compared with experimental data collected in five years time, obtaining perfect agreement, within the experimental parameters uncertainty.

Clearly, the legacy of the Brownian motion, which describes the most diverse equilibrium phenomena, such as the random motion of pollen in water and the microscopic current fluctuations of shot noise, extends to the case of nonequilibrium systems. Much research continues to be performed within this realm, the framework of stochastic thermodynamics being one instance of that [15].

## 1.5 Fluctuations Dissipation Relations

The study of the Brownian motion is important because its range of applicability includes an incredibly wide set of fluctuating phenomena. At the same time it is limited to equilibrium phenomena characterized by equipartition and by fast decay of correlations, which excludes, for instance, the gravitational antenna described above. To treat different phenomena, one may begin relaxing the condition of white spectrum on the stochastic term  $\Gamma$ , introducing retarded frictions. In particular one may consider

$$\dot{v}(t) = - \int_{-\infty}^t \gamma(t-t')v(t')\mathbf{d}t' + \Gamma(t), \quad \mathbb{E}[\Gamma(t)] = 0 \quad (1.138)$$

which is linear and can be treated by harmonic analysis using the Fourier integrals:

$$v(t) = \int_{-\infty}^{\infty} \tilde{v}(\omega)e^{i\omega t}\mathbf{d}\omega; \quad \Gamma(t) = \int_{-\infty}^{\infty} \tilde{\Gamma}(\omega)e^{i\omega t}\mathbf{d}\omega; \quad \gamma(t) = \int_{-\infty}^{\infty} \tilde{\gamma}(\omega)e^{i\omega t}\mathbf{d}\omega \quad (1.139)$$

where:

$$\tilde{v}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} v(t)e^{-i\omega t}\mathbf{d}t; \quad \tilde{\Gamma}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Gamma(t)e^{-i\omega t}\mathbf{d}t; \quad \tilde{\gamma}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \gamma(t)e^{-i\omega t}\mathbf{d}t \quad (1.140)$$

Now, causality implies that times  $t' > t$  have no influence on the process developed up to time  $t$ , hence the friction term must obey:

$$\gamma(t) = \begin{cases} \gamma(t) & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases} \quad (1.141)$$

Therefore, one can write:

$$\int_{-\infty}^t \gamma(t-t')v(t')\mathbf{d}t' = \int_{-\infty}^{\infty} \gamma(t-t')v(t')\mathbf{d}t' \quad (1.142)$$

as a convolution integral. Substituting Eq.(1.139) in Eq.(1.138) leads to:

$$i\omega\tilde{v}(\omega) = -\tilde{\gamma}(\omega)\tilde{v}(\omega) + \tilde{\Gamma}(\omega), \quad (1.143)$$

hence

$$\tilde{v}(\omega) = \frac{\tilde{\Gamma}(\omega)}{i\omega + \tilde{\gamma}(\omega)} \quad \text{and} \quad I_v(\omega) = \frac{I_{\Gamma}(\omega)}{|i\omega + \tilde{\gamma}(\omega)|^2} \quad (1.144)$$

At this stage, many possible routes can be taken, because different relations between  $\gamma$  and  $\Gamma$  can be assumed. However, given the stunning success of the Brownian motion theory, where it applies, it seems necessary to develop a framework that reduces to that theory, when conditions return to those of the Brownian motion. Moreover, one may wish that the autocorrelation of  $v$ , that is derived from the spectrum of  $\Gamma$ , corresponds to a thermodynamic equilibrium state. That imposes conditions on  $I_{\Gamma}$ . For instance, expressing the retarded viscosity in terms of its real and imaginary parts,  $\tilde{\gamma} = Re[\tilde{\gamma}] + iIm[\tilde{\gamma}]$ , one may require:

$$I_\Gamma(\omega) = \frac{k_B T}{m\pi} \operatorname{Re} [\tilde{\gamma}(\omega)], \quad \operatorname{Re} [\tilde{\gamma}(\omega)] \geq 0 \quad (1.145)$$

which mimics Eq. (1.100). Here, we require  $\operatorname{Re} [\tilde{\gamma}(\omega)] \geq 0$ , because the power spectrum cannot be negative. To obtain this result, one may extend the definition of  $\gamma$  to negative times, setting  $\hat{\gamma}(-t) = \hat{\gamma}(t) = \gamma(t)$ , for  $t \geq 0$ , and then try the following assumption:

$$\mathbb{E}[\Gamma(t_1)\Gamma(t_2)] = \frac{k_B T}{m} \hat{\gamma}(t_1 - t_2) \quad (1.146)$$

Considering that  $\tilde{\gamma}(\omega) = \int_0^\infty dt \gamma(t) \exp(-i\omega t)/2\pi$ , the result is:

$$\begin{aligned} I_\Gamma(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbb{E}[\Gamma(0)\Gamma(t)] e^{-i\omega t} dt = \frac{k_B T}{2\pi m} \left[ \int_{-\infty}^0 \hat{\gamma}(t) e^{-i\omega t} dt + \int_0^\infty \hat{\gamma}(t) e^{-i\omega t} dt \right] \\ &= \frac{k_B T}{2\pi m} [\tilde{\gamma}(-\omega) + \tilde{\gamma}(\omega)] = \frac{k_B T}{m\pi} \operatorname{Re} [\tilde{\gamma}(\omega)] \end{aligned} \quad (1.147)$$

where the last equality holds if  $\tilde{\gamma}(-\omega) = \tilde{\gamma}(\omega)^*$ . Also, taking  $\gamma(t) = \gamma\delta(t)$ , with  $\gamma \in \mathbb{R}$ , one falls back in the original situation with  $\delta$ -correlated noise and obtains:

$$I_\Gamma = \frac{\gamma k_B T}{\pi m} \quad (1.148)$$

To generalize the mobility, observe first that it can be cast in the following form:

$$\mu = \frac{D}{k_B T} = \lim_{t \rightarrow \infty} \frac{\mathbb{E}[(x(t)-x(0))^2]}{2tk_B T} = \lim_{t \rightarrow \infty} \frac{1}{2tk_B T} \int_0^t dt_1 \int_0^t dt_2 \mathbb{E}[v(t_1)v(t_2)] \quad (1.149)$$

$$= \lim_{t \rightarrow \infty} \frac{1}{2tk_B T} \int_0^t dt_1 \int_0^t dt_2 \mathbb{E}[v(0)v(t_2 - t_1)] = \lim_{t \rightarrow \infty} \frac{1}{k_B T t} \int_0^t (t-s) \mathbb{E}[v(0)v(s)] ds \quad (1.150)$$

$$= \frac{1}{k_B T} \lim_{t \rightarrow \infty} \int_0^t \mathbb{E}[v(0)v(s)] ds \quad (1.151)$$

where the first term of line (1.150) holds because we consider a steady state, while the second term of line (1.150) and line (1.151) can be derived with a little bit of algebra, provided the following is satisfied:

$$\left| \int_0^\infty \phi_v(s) ds \right| < \infty, \quad \text{and} \quad \int_0^t s \phi_v(s) ds = o(t) \text{ for } t \rightarrow \infty \quad (1.152)$$

To illustrate what this means, consider a positive  $\phi_v$ , although in general it fluctuates between positive and negative values. In that case, the second condition in Eq. (1.152)

requires  $\phi_v(t)$  to decay in time faster than  $t^{-2}$ . This constraint is sufficient, not necessary; it remains that a sufficiently fast correlations decay is required for the transport coefficients to exist.

The expression (1.151) of  $\mu$  is one example of Green-Kubo integral; there is one such Green-Kubo integral for every linear transport coefficient, obtained from the autocorrelation of the current of interest. A transport process is called *normal* if that autocorrelation decays sufficiently fast in time, that the integral, hence the transport coefficient, exists.

Observing that the integral (1.151) is a kind of Fourier transform at zero frequency, one may adopt the following expression for the frequency dependent mobility:

$$\mu(\omega) = \frac{1}{k_B T} \int_0^{\infty} dt \mathbb{E} [v(0)v(t)] e^{-i\omega t}, \quad (1.153)$$

so that one may also define the frequency dependent diffusion coefficient as  $D(\omega) = k_B T \mu(\omega)$ . Now, note that in a stationary state, time translation invariance holds, therefore one may write:

$$\phi_v(t) = \mathbb{E} [v(0)v(t)] = \mathbb{E} [v(-t)v(0)] = \phi_v(-t) \quad (1.154)$$

which implies:

$$I_v(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_v(t) e^{-i\omega t} dt = \frac{1}{2\pi} \int_{-\infty}^0 \phi_v(t) e^{-i\omega t} dt + \frac{1}{2\pi} \int_0^{\infty} \phi_v(t) e^{-i\omega t} dt \quad (1.155)$$

$$= \frac{k_B T}{2\pi} [\mu(\omega) + \mu(-\omega)] = \frac{k_B T}{\pi} \text{Re} [\mu(\omega)] \quad (1.156)$$

where the last equality holds if  $\mu(-\omega) = \mu(\omega)^*$ . Equation (1.156) is consistent with (1.37), for which  $\mu = 1/m\gamma$ , since (1.144) and (1.156) imply  $\mu(\omega) = \tilde{\gamma}(\omega)/m |i\omega + \tilde{\gamma}(\omega)|^2$ , hence  $\mu(0) = 1/m\tilde{\gamma}(0)$ .

A second way to compute  $I_v$ , derives from Eqs. (1.144) and (1.145):

$$I_v(\omega) = \frac{I_{\Gamma}(\omega)}{|i\omega + \tilde{\gamma}(\omega)|^2} = \frac{k_B T}{m\pi} \frac{\text{Re}[\tilde{\gamma}(\omega)]}{|i\omega + \tilde{\gamma}(\omega)|^2} = \frac{k_B T}{m\pi} \frac{1}{2} \left[ \frac{1}{i\omega + \tilde{\gamma}(\omega)} + \frac{1}{-i\omega + \tilde{\gamma}(\omega)^*} \right] \quad (1.157)$$

Comparing Eq. (1.156) with Eq. (1.157), and using Eq. (1.153), the mobility writes:

$$\mu(\omega) = \frac{1}{m [i\omega + \tilde{\gamma}(\omega)]} = \frac{1}{k_B T} \int_0^{\infty} dt \mathbb{E} [v(0)v(t)] e^{-i\omega t} \quad (1.158)$$

Analogously, Eq. (1.147) the friction coefficient takes the form:

$$\tilde{\gamma}(\omega) = \frac{m}{k_B T} \int_0^{\infty} dt e^{-i\omega t} \mathbb{E} [\Gamma(0)\Gamma(t)] \quad (1.159)$$

Equations (1.158) and (1.159) are called a *Fluctuation-Dissipation Relation* (FDR) of first kind and of second kind, respectively. The first kind gives the complex mobility (admittance, in general) in terms of the autocorrelation of the velocity (flow, in general). The second kind gives the complex viscosity (impedance, in general) in terms of the autocorrelation of the the random force. As explained in Ref. [16], these two kinds of FDR imply that the response of a system to external actions which perturb its equilibrium is linked to the spontaneous thermal fluctuations in absence of perturbing forces. The FDR of the first kind has to be considered more fundamental than the second, since it refers to experimentally accessible quantities (the flows), while the second kind relies on the distinction between frictional and random forces, which is made problematic by Eq. (1.146). Other popular Green-Kubo integrals are the following:

$$D_i = \frac{1}{3} \int_0^{\infty} dt \mathbb{E} [v_i(0)v_i(t)], \quad \text{self diffusion in 3D} \quad (1.160)$$

$$\lambda = \frac{V}{3k_B T^2} \int_0^{\infty} dt \mathbb{E} [J_q(0)J_q(t)], \quad \text{thermal conductivity} \quad (1.161)$$

$$\eta = \frac{V}{3k_B T} \int_0^{\infty} dt \mathbb{E} [P_{xy}(0)P_{xy}(t)], \quad \text{shear viscosity} \quad (1.162)$$

where  $V$  is the volume occupied by the system of interest and  $T$  is its temperature. It should be noted that these formulae are obtained equally the same from stochastic as well as deterministic processes.

To conclude this section, it is interesting to analyze the procedure we have followed: given the result we wanted to obtain (consistently generalize the Brownian motion theory), we have searched for the conditions that produce it. Consequently, the model we have constructed certainly yields the desired result. The question is now whether systems of physical interest satisfy the imposed conditions actually exist. This is standard practice in physics, often more useful than the straight logical deductions from general principles, which are frequently cumbersome or even impossible. For instance, in statistical physics, one typically relies on macroscopic observations to infer the form of the molecular interaction potentials, not vice versa. The theory of Brownian motion is of the other kind: some general ideas on the microscopic dynamics led to predictions on the macroscopic behaviour, which were subsequently experimentally verified.

## 1.6 Particle Systems in Phase Space

Let us now relate the microscopic atomic description of matter to the macroscopic thermodynamic description. The first question, in view of the observations made in the Introduction, is which microscopic model should one use for such a task. A priori it is not possible to tell, therefore we begin with classic Hamiltonian mechanics. Why? Because it has proven extremely successful in describing a huge variety of phenomena that occur at the scales of our daily life, and well beyond, reaching astrophysical objects. This, at least, was the opinion expressed by Laplace, when stating [17]:

We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.<sup>10</sup>

Now, rather than describing a moderate number of macroscopic objects, for which classical mechanics has proven so successful, we want to describe an exceedingly large assembly of very small objects. This may be a cause of concern; as usual, we will judge the validity of this approach from the results it will produce.

Given a system of  $N$  particles each having  $d$  degrees of freedom, we introduce the space of “microscopic phases”, or simply phase space,  $\mathcal{M} \subset \mathbb{R}^{2dN}$ , and its points  $\Gamma = (\mathbf{q}, \mathbf{p}) \in \mathcal{M}$ , where  $\mathbf{q}$  represents the  $dN$ -dimensional configuration vector, and  $\mathbf{p}$  the  $dN$ -dimensional vector of momenta. It is assumed that each  $\Gamma$  fully represents the microscopic state of the particle system. Then, in the case of Hamiltonian dynamics, there exists an energy function  $H = H(\Gamma)$ , such that the time evolution is given by:

$$\dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}, \quad \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}} \quad (1.163)$$

We denote by  $S^t : \mathcal{M} \rightarrow \mathcal{M}$  the evolution operator for a time  $t \in \mathbb{R}$ , meaning that  $S^t \Gamma \in \mathcal{M}$  is the microstate at time  $t$ , if it was  $\Gamma$  at time 0.

To connect with the macroscopic level of observation, one introduces the notion of phase function,  $\mathcal{O} : \mathcal{M} \mapsto \mathbb{R}$ , and, considering that a measurement takes a positive time, one assumes that it yields a time average of one phase function, also called *observable*. If the measurement takes a time  $\tau$ , and it starts when the microstate is  $\Gamma$ , the result of the measurement is:

$$\overline{\mathcal{O}}_{0,\tau}(\Gamma) = \frac{1}{\tau} \int_0^\tau \mathcal{O}(S^t \Gamma) dt \quad (1.164)$$

---

<sup>10</sup>Note, however, that considering that such an intellect is out of our reach, the best approach to nature is probabilistic.



If one accepts this picture, it follows that measurements depend on  $\tau$ , hence are subjective because different observers may choose different  $\tau$ 's, and on  $\Gamma$ , hence their result is a stochastic variable, because  $\Gamma$  is unknown. However, this contradicts Thermodynamics, that is an objective and deterministic theory, universally confirmed by experimental tests. One possibility to accord Eq. (1.164) with experience is to assume that  $\tau$  is very large, virtually infinitely larger than the microscopic scales concerning the evolution of  $\Gamma$ , and that during such a long time,  $\mathcal{O}$  has explored many times its range of values, with proper frequencies, so that the initial condition  $\Gamma$  is irrelevant. In that case, one may indeed write:

$$\frac{1}{\tau} \int_0^{\tau} \mathcal{O}(S^t \Gamma) dt \approx \overline{\mathcal{O}}(\Gamma) = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^{\tau} \mathcal{O}(S^t \Gamma) dt \approx o \in \mathbb{R} \quad (1.165)$$

where mathematically the first approximate equality is due to the fact  $\tau$  can be very large, but does not need to be infinite, and the second to the fact the range of  $\mathcal{O}$  may not be perfectly explored. As long as the approximations fall below the scale of thermodynamic interest, equality can thus be used, and  $o$  represents the result of a measurement. This picture is justified by Fermi as follows [18]:

Studying the thermodynamical state of a homogeneous fluid of given volume at given temperature [...] we observe that there is an infinite number of states of molecular motion that correspond to it. With increasing time, the system exists successively in all the dynamical states that correspond to the given thermodynamical state. From this point of view we may say that a thermodynamical state is the ensemble of all the dynamical states through which, as a result of the molecular motion, the system is rapidly passing.

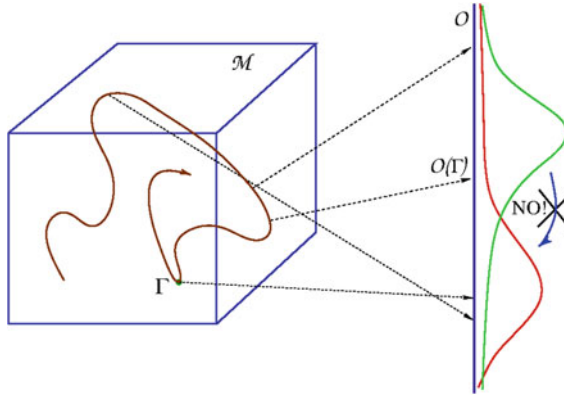
One should note the term “rapidly” is not guaranteed in general.<sup>11</sup> However, when the microscopic values of  $\mathcal{O}$  are indeed sufficiently rapidly explored, compared to macroscopic observation times, a single system of interest reveals itself through  $\mathcal{O}$  as the average over the *ensemble* of such possibilities, cf. Fig.1.8. To the observer, the system revealed by  $\mathcal{O}$  *appears* like the average over the ensemble of all its microscopic phases, suggesting that the result of a measurement may be obtained computing an average with respect to a probability distribution on phase space, called *ensemble*.

**Remark 1.6.1** This requires the macroscopic state to be stationary; if it shifts during the measurement, the observable cannot explore its range with the frequencies corresponding to that state, and different initial microstates may lead to different observable values. In fact, this is the case of e.g. ageing systems.

Given a dynamical system like the above Eqs.(1.163), it is not obvious that Eq.(1.165) holds, and even if it does, the statement is so complex that it may be impossible to prove. Therefore, one introduces the so-called *Ergodic Hypothesis*, commonly stating that  $\mathcal{M}$  is densely explored by almost all trajectories. The result

---

<sup>11</sup>Think e.g. of ageing systems.



**Fig. 1.8** While the phase wanders in phase space  $\mathcal{M}$ , the phase function  $\mathcal{O}$  explores its range of values. Given the ultra-astronomically large dimensionality of  $\mathcal{M}$ , it would take unrealistically long times to finely explore such a space. The range of  $\mathcal{O}$  may nevertheless be thoroughly spanned, allowing Eq. (1.165) to hold. However, the distribution of  $\mathcal{O}$ -values must be stationary, lest it is not properly sampled in time, and averages depend on the initial condition

can be formulated as follows: there is an invariant probability distribution  $\mu$  on  $\mathcal{M}$ , such that a measurement yields:

$$\overline{\mathcal{O}}(\Gamma) = \int \mathcal{O}(\Gamma) d\mu(\Gamma) \equiv \langle \mathcal{O} \rangle_{\mu}, \quad \text{for } \mu\text{-almost every } \Gamma \in \mathcal{M} \quad (1.166)$$

for every phase function  $\mathcal{O}$ . This suffices for  $\mathcal{O}(S^t\Gamma)$  to span its range of values, but it is way too strong a condition to hope it holds as stated, which mathematically amounts to the condition of *indecomposability*.<sup>12</sup> Nevertheless, experience has shown that this hypothesis works very well in describing equilibrium systems, when  $\mu$  is one of the classical ensembles: microcanonical, for isolated systems; canonical, for systems in equilibrium with a heat bath, and grand-canonical, for systems in equilibrium with heat and particle reservoirs.

This fact can be explained as follows. Through a single variable  $\mathcal{O}$ , the system may appear like a phase space average even if its phase space trajectory  $S^t\Gamma$  does not explore finely  $\mathcal{M}$ . In fact, as long as the range of values of  $\mathcal{O}$  has been experienced with proper frequencies, the result is the same, cf. Fig. 1.8. If one observes more than one phase variable, it may take longer for both to have explored their respective

<sup>12</sup> What can actually be stated in general is much less and physically scarcely interesting. In practice, in general terms, one can prove that time averages exist with probability 1 with respect to the steady state distribution. In a dissipative case, this probability is piled up in set of zero phase space volume, hence it concerns too limited a fraction of the interesting possible initial conditions for the system of interest. Moreover, rather than the equality of the phase space average with the time averages, one obtains the equality of the phase space averages with the phase space average of the time averages. This is the substance of the Birkhoff ergodic theorem, see e.g. Ref. [19]. To obtain more is a rather hard task, that can be performed focussing on special cases [20].

ranges. If this condition had to be verified by all possible phase functions, the time required for a systems of many degrees of freedom could be so hugely long, to make the ergodic hypothesis physically irrelevant. Luckily, macroscopic observations only concern a handful of variables, which in addition are quite well behaved. This explains why ergodicity is continually and successfully adopted.

In some cases, this reasoning can be rigorously justified. The approach developed by Khinchin for rarefied gases does that, explaining that phase space subtleties are irrelevant, compared to the fact that [21]:

- (a) macroscopic systems are made of very many particles:  $N \gg 1$ ;
- (b) only several and special phase functions are physically relevant;
- (c) it does not matter if ensemble averages disagree with time averages on a limited sets of trajectories.

For rarefied gases, the relevant phase functions are sums of molecular contributions,  $f(\Gamma) = \sum_{n=1}^N f_n(\mathbf{q}_n, \mathbf{p}_n)$ , where  $(\mathbf{q}_n, \mathbf{p}_n)$  is the vector of configurations and momentum of the  $n$ -th particle. These functions are appropriate for the pressure, temperature and density of rarefied gases, whose energy can also be expressed as  $H = \sum_{n=1}^N H_n(\mathbf{q}_n, \mathbf{p}_n)$ , because interactions among particles are energetically negligible.<sup>13</sup> Then, in the microcanonical case, which in this case means a uniform probability distribution in  $\mathcal{M}$ , Khinchin proved the validity of the following relation:

$$\text{Prob} \left( \frac{|\bar{f} - \langle f \rangle|}{|\langle f \rangle|} \geq K_1 N^{-1/4} \right) \leq K_2 N^{-1/4}, \quad (1.167)$$

where  $K_1$  and  $K_2$  are positive constants,  $\bar{f}$  is the time average of  $f$ , and  $\langle f \rangle$  its phase space average in the microcanonical ensemble. In other words, the probability (in the microcanonical sense) that time averages differ by a small amount from the phase space averages is small if  $N$  is large: the larger  $N$  the smaller the probability of even smaller differences.

In this framework, the physically relevant ergodicity follows by and large from the  $N \gg 1$  condition, combined with the validity of the law of large numbers, which make the sum variables *practically constant*, and the exploration of their range fast. The details of the microscopic dynamics, including transitivity or lack of transitivity result irrelevant, while ensembles, i.e. probabilities in phase space turn considerably useful; even though probability, *per se* is an immaterial and abstract mathematical notion, it becomes “real”, in some sense, under the above conditions.

To understand more deeply when and why probability can be treated as real, which arguably makes it the most useful mathematical tool in Physics, let us consider evolving states, rather than stationary states, thus challenging Remark 1.6.1. If a stationary probability distribution correctly represents an equilibrium state, an evolving probability might, perhaps under different conditions, represent an evolving state.

---

<sup>13</sup>Interactions are however essential for the condition of LTE, hence for the existence of thermodynamic properties, to be established.

This requires a rule governing time dependent probabilities in phase space. What rule? If probability has to adhere to the behavior of material objects, its evolution should presumably be related to that of matter, and matter moves according to the laws of mechanics. However, unlike material objects, a chunk of probability does not have inertia, and there are no forces that push it like Newton's second law prescribes; even the space in which probability may somehow move is quite peculiar compared to the three-dimensional space where material objects move. The only property of mass that probability seems compelled to share is that it should be conserved.<sup>14</sup> One may then argue that probability moves in phase space like a fluid moves in real space; in other words, one may assume that sets of phase space points –that move in phase space according to dynamical laws such as Eq. (1.163)– carry probability with themselves, like fluid elements carry mass with themselves. To formalize this idea, let us concisely write the (not necessarily Hamiltonian) evolution equation for phases as:

$$\dot{\Gamma} = G(\Gamma), \quad \Gamma \in \mathcal{M} \quad (1.168)$$

where  $G : \mathcal{M} \rightarrow \mathcal{M}$  is a vector field, and let us endow  $\mathcal{M}$  with a probability density  $f$ . The above assumption amounts to postulating that  $f$  obeys a continuity equation in  $\mathcal{M}$ :

$$\frac{\partial f}{\partial t} = -\nabla_{\Gamma} \cdot (f G) = f \Omega^f; \quad \text{or} \quad \frac{df}{dt} = -f \Lambda \quad (1.169)$$

where

$$\Lambda(\Gamma) = \nabla_{\Gamma} \cdot G|_{\Gamma} \quad (1.170)$$

is the phase space volume variation rate at  $\Gamma$ , and

$$\Omega^f(\Gamma) = -G(\Gamma) \cdot \nabla_{\Gamma} \ln f|_{\Gamma} - \Lambda(\Gamma) \quad (1.171)$$

is known as *dissipation function*. Equation (1.169) is the generalized Liouville equation, which reduces to

$$\frac{df}{dt} = 0 \quad (1.172)$$

in the case of Hamiltonian dynamics, because:

$$\Lambda(\Gamma) = \sum_i \left[ \frac{\partial}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial}{\partial p_i} \frac{\partial H}{\partial q_i} \right] = 0$$

Denoting the time integrals and averages of a phase function  $\mathcal{O}$  by:

---

<sup>14</sup>For instance, what would it mean that the probability of the set containing all possible events turns larger than 1? Or that it gets smaller than 1? It would indicate that at some point one was not or is not really treating all possible events.

$$\mathcal{O}_{s,t}(\Gamma) = \int_s^t \mathcal{O}(S^u \Gamma) \mathbf{d}u, \quad \overline{\mathcal{O}}_{s,t} = \frac{1}{t-s} \int_s^t \mathcal{O}(S^u \Gamma) \mathbf{d}u \quad (1.173)$$

the solution of Eq. (1.169) can be formally written as:

$$\begin{aligned} f^{(t)}(\Gamma) &= \exp \{ -\Lambda_{-t,0}(\Gamma) \} f^{(0)}(S^{-t} \Gamma) \\ &= \exp \{ \Omega_{-t,0}^{f^{(0)}}(\Gamma) \} f^{(0)}(\Gamma) \end{aligned} \quad (1.174)$$

Then, a density  $f^{(0)}$  evolves, unless  $\Omega_{-t,0}^{f^{(0)}}$  identically vanishes. Thanks to these expressions, one may compute the time dependent phase space averages of phase functions, obtaining:

$$\langle \mathcal{O} \rangle_{\sqcup} = \int_{\mathcal{M}} \mathcal{O}(\Gamma) \{^{(\sqcup)}(\Gamma) \mathbf{d}\Gamma = \int_{\mathcal{M}} \mathcal{O}(\Gamma) \{^{(t)}(S^{-\sqcup} \Gamma) \mathcal{J}^{-\sqcup}(\Gamma) \mathbf{d}\Gamma \quad (1.175)$$

where  $J^{-t}(\Gamma) = \exp \{ -\Lambda_{-t,0}(\Gamma) \}$ . Then, introducing  $\Gamma = S^t Y$  we get:

$$\int_{\mathcal{M}} \mathcal{O}(S^t Y) f_0(Y) J^{-t}(S^t Y) J^t(Y) \mathbf{d}Y \quad (1.176)$$

where  $J^t(Y)$  is the variation of phase space volumes along a trajectory starting at  $Y$  and proceeding for a time  $t$ , and  $J^{-t}(S^t Y)$  is the variation of the backward evolution coming back from  $S^t Y$ , which can be easily shown to obey  $J^{-t}(S^t Y) = 1/J^t(Y)$ .<sup>15</sup> One then gets:

$$\langle \mathcal{O} \rangle_{\sqcup} = \int_{\mathcal{M}} \mathcal{O}(\Gamma) \{^{(\sqcup)}(\Gamma) \mathbf{d}\Gamma = \int_{\mathcal{M}} \mathcal{O}(S^{\sqcup} \Gamma) \{^{(t)}(\Gamma) \mathbf{d}\Gamma = \langle \mathcal{O} \circ S^{\sqcup} \rangle, \quad (1.177)$$

proving that evolution of observables can equivalently be obtained evolving either the microscopic quantity itself or the probability density. This is the classical analog of the equivalence of the Heisenberg and Schrödinger pictures of quantum mechanics.

### 1.6.1 Boltzmann Equation and H-Theorem

Let us specify the above general treatment of probabilities in phase space, to the case of  $N$  identical spherical hard particles of mass 1, that are subjected to no external forces, and that only interact elastically when they collide. We may think of their

---

<sup>15</sup>Note,  $S^{-t}$  traces backward the phase space trajectory; it is not the time reversed trajectory, cf. (1.195) below.

interaction potential as being 0 when the particles centers are at a distance larger than their diameter, and infinitely high when this distance equals their diameter, so that particles are neither deformed, nor they penetrate each other, when they collide. Denote by

$$f_N^{(N)}(\Gamma) = f_N^{(N)}(\mathbf{q}_1, \mathbf{p}_1; \dots; \mathbf{q}_N, \mathbf{p}_N; t) \quad (1.178)$$

an initial density in the phase space  $\mathcal{M}$ , with notation indicating that the joint probability density of  $N$  particles out of  $N$  is considered. In absence of external forces, there are no accelerations between collisions, hence  $\dot{\mathbf{p}}_i = 0$  for every particle, and the Liouville equation can be written as:

$$\frac{\partial f_N^{(N)}}{\partial t} + \left( \frac{\partial}{\partial \Gamma} \cdot \dot{\Gamma} + \dot{\Gamma} \cdot \frac{\partial}{\partial \Gamma} \right) f_N^{(N)} = \frac{\partial f_N^{(N)}}{\partial t} + \sum_{i=1}^N \mathbf{p}_i \cdot \frac{\partial f_N^{(N)}}{\partial \mathbf{q}_i} = 0 \quad (1.179)$$

where we have used the fact that the dynamics is Hamiltonian, hence the first term in brackets is null, and that  $\dot{\Gamma} = (\dot{\mathbf{q}}_1, 0; \dots; \dot{\mathbf{q}}_N, 0)$ . If we introduce the  $s$ -particle distribution function:

$$f_N^{(s)}(\mathbf{q}_1, \mathbf{p}_1; \dots; \mathbf{q}_s, \mathbf{p}_s; t) = \int d\mathbf{q}_{s+1} d\mathbf{p}_{s+1} \dots d\mathbf{q}_N d\mathbf{p}_N f_N^{(N)}(\mathbf{q}_1, \mathbf{p}_1; \dots; \mathbf{q}_N, \mathbf{p}_N; t) \quad (1.180)$$

and we integrate the Liouville equation over the variables  $\mathbf{q}_{s+1}, \mathbf{p}_{s+1} \dots \mathbf{q}_N, \mathbf{p}_N$ , we obtain [22]:

$$\frac{\partial f_N^{(s)}}{\partial t} + \sum_{i=1}^s \mathbf{p}_i \cdot \frac{\partial f_N^{(s)}}{\partial \mathbf{q}_i} = F^{(s)}(f_N^{(s+1)}, f_N^{(s+1)'}) \quad (1.181)$$

where  $F$  is a function of the joint  $(s+1)$  particles distribution before collision  $f_N^{(s+1)'}$ , and of the corresponding distribution after collision  $f_N^{(s+1)}$ . The important fact is that (1.181) is not a closed equation: computing  $f_N^{(s)}$  requires knowledge of  $f_N^{(s+1)}$ . In particular, taking  $s=1$  and making  $F^{(1)}$  explicit by solving the elastic collision dynamics, one obtains:

$$\frac{\partial f_N^{(1)}}{\partial t} + \mathbf{v} \cdot \frac{\partial f_N^{(1)}}{\partial \mathbf{q}} = (N-1)\sigma^2 \int [f_N^{(2)'} - f_N^{(2)}] |\mathbf{v} \cdot \mathbf{n}| d\mathbf{p}_* d\mathbf{n} \quad (1.182)$$

for the 1-particle probability distribution function, the result of projecting out all particles but one, from the phase space probability distribution function  $f_N^{(N)}$ . Here,  $\sigma^2$  is the collision cross section,  $\mathbf{v} = \mathbf{p}$  because the mass of particles is 1,  $\mathbf{n}$  is the unit vector in the direction joining the centers of the two particles concerning  $f_N^{(2)}$ , the distribution after collision, and  $f_N^{(2)'}$ , the distribution before collision. In turn  $\mathbf{p}_*$  is the momentum of the second particle.

In order to solve this equation, one needs an assumption on  $f_N^{(2)}$ , because is not possible to solve the whole of equations for the  $f_N^{(s)}$ . Boltzmann proposed the following closure hypothesis:

$$f_N^{(2)}(q_1, p_1; q_*, p_*; t) = f_N^{(1)}(q_1, p_1; t) f_N^{(1)}(q_*, p_*; t) \quad (1.183)$$

which is known in the kinetic theory of gases as the *stosszahlansatz*, or hypothesis of molecular chaos. Indeed, this assumption amounts to state that when two particles collide, they are independent. As the dynamics of particles is deterministic, this statement may only have a statistical meaning, and requires some kind of *randomness*, that can be legitimately called *chaos*.<sup>16</sup> This makes sense for systems with many particles, two of which collide at time  $t$ , since they may have hardly interacted before, and can be considered independent. But clearly this independence does not hold for particles which have just collided. Moreover, independence is harder to achieve for  $s$  particles if  $s$  is larger, because  $s$  particles occupy a volume of order  $O(s\sigma^3)$  and particles cannot overlap. Therefore,  $s = 1$  is the best candidate for the independence of particles. Furthermore, for  $s = 1$ , the limit  $N \rightarrow \infty$ ,  $\sigma \rightarrow 0$  should foster the validity of the *stosszahlansatz*, when the total cross section for collision does not vanish, and a randomizing mechanism is in place. Grad identified the scaling regime in which this makes sense, now known as Boltzmann-Grad limit [23]: it consists in keeping  $N\sigma^2$  positive and finite, while  $N$  grows, so that  $N\sigma^3 \rightarrow 0$ :

$$N\sigma^2 = \text{constant} > 0 \implies \sigma^2 \sim \frac{1}{N}, \quad N\sigma^3 \sim N^{-\frac{1}{2}} \quad (1.184)$$

This way, there is a net effect coming from particles collisions, necessary to randomize the motion and, at the same time, the excluded volume that may lead to correlations vanishes, while  $N$  can be as large as desired. Physically, this picture corresponds to a rarefied gas, in which particles collide, but their interaction energy is negligible compared to their kinetic energy. The resulting equation:

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{q}} = (N-1)\sigma^2 \int [f^{(1)'} f_*^{(1)'} - f(1) f_*^{(1)}] |\mathbf{v} \cdot \mathbf{n}| d\mathbf{p}_* d\mathbf{n} \quad (1.185)$$

is the celebrated Boltzmann equation, where we have simplified notation writing  $f$  in place of  $f_N^{(1)}$ . Its applicability goes well beyond the bounds of its strict derivation, which is that of rarefied gases in regular containers. For instance, introducing external electric potentials, it is applied to transport of electrons in solids [24]; adding nuclear cross sections, it is applied to transport of neutrons in conventional nuclear reactors, or to cold as well as hot nuclear fusion technology [25, 26]; in linearized and/or discretized versions, such as those known as lattice Boltzmann models, it is applied to a great variety of fluids, including blood cells in blood vessels [27].

An important result concerning Eq. (1.185) is that  $f_N^{(2)}$  remains factorized in time, if it is such at start, keeping the validity of the model in time. More precisely, assuming that  $f_N^{(s)}$  exists and is well behaved for any fixed  $s$ , when  $N \rightarrow \infty$  and  $N \rightarrow \sigma^2$

---

<sup>16</sup>Note: in the theory of dynamical systems, the term chaos is often used to indicate systems that have at least on positive Lyapunov exponent. This notion was not available to Boltzmann and, indeed, he did not need this notion of chaos.

constant, one obtains that  $f_N^{(s)}$  remains factorized in time, as product of a number  $s$  of  $f_N^{(1)}$  factors, if it is so factorized at the beginning [22].

Given the  $\mathcal{H}$ -functional defined by:

$$\mathcal{H} = \int f \log f \, d\mathbf{p} \quad (1.186)$$

where  $f$  is a 1-particle distribution, one of the main conceptual results stemming from the Boltzmann equation is the  $\mathcal{H}$ -theorem. Provided  $f$  is the solution of the Boltzmann equation under boundary conditions of regular reflecting walls, hence for an isolated system, the theorem states:

$$\frac{d}{dt} \int \mathcal{H}(\mathbf{q}) \, d\mathbf{q} \leq 0 \quad (1.187)$$

where equality only holds when  $f$  is the Maxwell-Boltzmann distribution.

For isolated dilute gases,  $\mathcal{H}$  reduces to the celebrated Boltzmann entropy

$$S_B = k_b \log W \quad (1.188)$$

that states that equal volumes in phase space correspond to events of equal probability, because the number  $W$  of different ways in which a given thermodynamic state can be microscopically realized is identified with a given volume in phase space. Equation (1.188) is also called *bridge law*, since it connects the microscopic description afforded by  $W$  with the macroscopic thermodynamic description provided by the entropy and its derivatives, via the amazing Boltzmann constant  $k_b$ . Also known as Boltzmann postulate, Eq. (1.188), was rewritten by Einstein as:

$$\text{Pr}(\text{state}) \propto e^{\Delta S/K_B} \quad (1.189)$$

meaning that a fluctuation away from the equilibrium state, characterized by a (negative) variation of entropy  $\Delta S$ , is exponentially unlikely. This begins the industry of fluctuation theories, and imparts momentum to stochastic modeling in Physics.

The distribution  $f$  appearing in the Boltzmann equation is a probability distribution obtained by projecting down the probability distribution for  $N$  particles on the  $2dN$  exceedingly high dimensional phase space, to the  $2d$  dimensional 1-particle space.<sup>17</sup> In other words,

$$f(\mathbf{q}, \mathbf{p}; t) d\mathbf{q} d\mathbf{p} \quad (1.190)$$

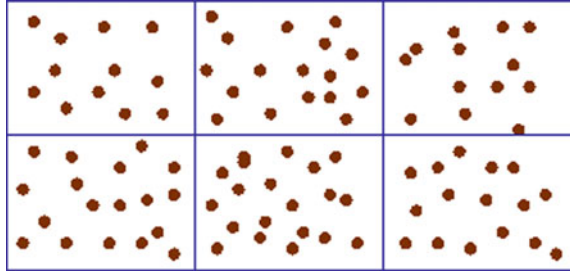
represents the probability of finding at time  $t$  one particle in the elementary volume  $d\mathbf{q} d\mathbf{p}$ , i.e. to find one particle within a volume of size  $d\mathbf{q}$  around  $\mathbf{q}$ , with momentum within a set of size  $d\mathbf{p}$  around  $\mathbf{p}$ . Furthermore, probability is intended here as the

---

<sup>17</sup>The  $2d$ -dimensional space of a single particle positions and velocities is called  $\mu$ -space, to distinguish it from the  $6N$  dimensional  $\Gamma$ -space, which is the phase space.



**Fig. 1.9** Subdivision of available physical and momentum space in discrete cells, sufficiently large to contain very many finite sized particles, but sufficiently small to look like a point to one observer



fraction of identical  $N$ -particle systems, the ensemble described by  $f_N^{(N)}$ , that have one particle within  $d\mathbf{q}d\mathbf{p}$ .

Of course, given a single system of interest, there is no need for any of its particles to actually lie within the volume  $d\mathbf{q}d\mathbf{p}$ , since there is a complementary fraction of the ensemble whose systems have no particles in  $d\mathbf{q}d\mathbf{p}$ . Therefore, one could conclude that even the Boltzmann equation is *immaterial*, and there is no compelling reason to take it as a description of a given material system. There is, however, a more physical derivation of the Boltzmann equation, that does not start from a continuous probability distribution on a very high dimensional continuous set of geometric points called phase space  $\mathcal{M}$ . This derivation considers a large, but finite number, of small, but finite size, particles with positions and velocities in a given volume  $d\mathbf{q}d\mathbf{p}$ . The distribution of these particles in such a volume is discrete, rather than continuous, and requires a proper coarse graining to be well represented by a continuous density of mass [28]. The construction requires the following steps.

Consider one system of  $N$  particles of unit mass, inside a given 3-dimensional container. It is a single macroscopic object, which has got nothing to do with collections of identical systems, or with probability distributions on abstract spaces. Both the container and the space of velocities are discretized in cells  $C_{ij}$  of size  $\Delta\mathbf{q}\Delta\mathbf{p}$  centered around a discrete set of lattice of points  $(\mathbf{q}_i, \mathbf{p}_j)$ , cf. Fig. 1.9. The cells must be sufficiently large that each of them contains a large number of particles  $n_{ij}$ , so that it makes sense to define the mass density  $\rho$  within them as the sum of the masses per unit volume:

$$n_{ij}(t) = \rho(\mathbf{q}_i, \mathbf{p}_j; t)\Delta\mathbf{q}\Delta\mathbf{p}; \quad \sum_{i,j} n_{ij} = N \quad (1.191)$$

Then, one may try to approximate this set of discrete mass densities  $\rho_{ij}$  with a continuous function  $\rho$ , so that

$$n_{ij}(t) \approx \int_{C_{ij}} \rho(\mathbf{q}, \mathbf{p}; t)d\mathbf{q}d\mathbf{p}; \quad N = \int_{\cup_{i,j} C_{ij}} \rho(\mathbf{q}, \mathbf{p}; t)d\mathbf{q}d\mathbf{p} \quad (1.192)$$

For this to make sense, cells must be sufficiently large to contain a large number of particles whose motion is sufficiently random, that those which enter or leave a cell are a negligible number compared to those inside. At the same time the cells must be sufficiently small compared to the observation scale, that they appear like a point. Clearly, this requires a wide separation of scales.

Suppose we now follow an elementary set  $\Delta\mathbf{q}\Delta\mathbf{p}$ , as the particles contained in it move in their container. Assume that  $\rho$  accurately describes the mass density in  $\mu$ -space and within the set under consideration. No particles enter or leave  $\Delta\mathbf{q}$ , even if its shape may be deformed in time. Particles may however enter or leave  $\Delta\mathbf{p}$  because colliding with each other they may suddenly fall into or exit from  $\Delta\mathbf{p}$ . Therefore, the total derivative may only change due to collisions, and one may write:

$$\frac{d\rho}{dt} = \frac{\partial\rho}{\partial t} + \mathbf{v} \cdot \frac{\partial\rho}{\partial\mathbf{q}} = \left(\frac{d\rho}{dt}\right)_{coll} \quad (1.193)$$

where the last term takes into account the gain or loss of particles from the  $\Delta\mathbf{p}$  volume, which in the absence of external forces is the only mechanism that allows velocities to change. This relation merely expresses the conservation of mass, and does not require any interpretation of probabilistic nature: it is based on the objective counting of particles (or measurement of mass density) and on their deterministic Hamiltonian dynamics. Assuming for a dilute gas that collisions of three or more particles are negligible, the collision term should be expressed as a function of the number of pairs of particles located inside  $\Delta\mathbf{q}$ , that may either enter (gain term) or leave (loss term)  $\Delta\mathbf{p}$  because of collisions. Let  $\rho^{(2)}$  be the density of such pairs. Strictly speaking one may proceed only solving the equations of motion of the  $N$  particles, once their initial condition is known. Since both are impossible, some statistical assumption on the collision term should be made, relying on both the large value of  $N$  and the disorder produced in configuration and velocity space by the collisions. Observing that Eq. (1.185) has same form as Eq. (1.193), one may formally adopt for  $\rho^{(2)}$  the stosszahlansatz:

$$\rho^{(2)}(\mathbf{q}, \mathbf{p}_1; \mathbf{q}, \mathbf{p}_2; t) = \rho^{(1)}(\mathbf{q}, \mathbf{p}_1) \rho^{(1)}(\mathbf{q}, \mathbf{p}_2) \quad (1.194)$$

This simply means that one particle may indifferently collide with the remaining  $N - 1$ , and that collisions of three or more particles at once give a negligible contribution. As a result, the equation for the mass distribution  $\rho$  turns identical to the one for the probability  $f$ , and its solution  $\rho$  enjoys the same properties enjoyed by  $f$ ,  $\mathcal{H}$ -theorem included.

The appropriateness of Eq. (1.194) follows from considerations similar to those for Eq. 1.183: under rarefied conditions, the number of molecular interactions per unit time and unit volume about  $\mathbf{q}$ , with momenta about  $\mathbf{p}_1$  and  $\mathbf{p}_2$ , can be proportional to the product of the densities of particles at that point, with the two different momenta, because each particle with momentum about  $\mathbf{p}_1$  may collide with each particle with momentum about  $\mathbf{p}_2$ . Whether the assumption applies or not, only experience can tell, like in the case of probabilities.

The conceptual difference between Eqs. (1.185) and (1.193) is enormous: in the first case, one considers the fraction of identical abstract systems that enjoy certain properties to the total of an ideal hypothetical continuum of identical systems; in the second case, one considers the mass of a single concrete, experimentally observable system. Clearly, there is no stringent reason for one quite sophisticated abstract notion to enjoy the same evolution of a materially touchable object. Nevertheless, this is precisely what happens if the assumptions in the derivations above are valid. That validity may well be too hard to prove within a mathematical framework, but experience has demonstrated that it does hold in very many situations. This is one case in which *probability* turns *material*: in some sense, probability can be identified with mass.

We noted that the structure of Eqs. (1.181) preserves the factorization in  $s$  single particle densities of the multiparticle densities, which is essential for the validity of the Boltzmann equation. Why should the distribution of pairs of particles be factorized to start with?

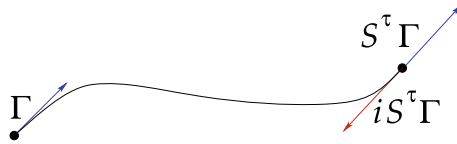
First of all, note that the construction of the Boltzmann equation proceeds from expressing the evolution of the density due to the collisions among particles. If this density is smooth at some point before two molecules collide, the question is whether collisions tend to preserve smoothness or to produce wrinkles and eventually singularities. In other words, whether collisions tend to maximize or reduce the distances among particles in  $(\mathbf{q}, \mathbf{p})$  space. Smoothness is indeed required for the equation to continue to hold in time, since derivatives have to exist. Therefore, because there are TRI dynamics that produce singularities as well as TRI dynamics that smooth out the distributions of particles both in phase space and in real space, we conclude that the Boltzmann equation is suitable only for the second case [22]. Then, repeated smoothing of the distribution gradually produces a homogenous state, which means maximum microscopic *disorder*. In phase space, this corresponds to the uniform distribution, known as the microcanonical ensemble, which amounts to lack of correlations and, ultimately to the factorization of the distributions.

In fact, collisions of hard spheres are defocussing, which seems to imply a wider volume occupied by the colliding particles, hence higher smoothness, after a collision. However, the situation is not that simple: Hamiltonian dynamics are time reversal invariant (TRI), which is to say that there exists an operator  $i : \mathcal{M} \rightarrow \mathcal{M}$ , that anticommutes with the time evolution:

$$S' i \Gamma = i S^{-1} \Gamma, \quad \text{and} \quad i i \Gamma = i^2 \Gamma = \Gamma, \quad \forall \Gamma \in \mathcal{M} \quad (1.195)$$

For instance, the operator defined by  $i(\mathbf{q}, \mathbf{p}) = (\mathbf{q}, -\mathbf{p})$  is the best known time reversal operation. Now, the problem is that reversibility makes possible a focusing collision for each defocusing collision ... (Fig. 1.10).

This is the basis of the so-called *reversibility objection* or *Loschmidt paradox*, which states that the  $\mathcal{H}$ -theorem cannot be a consequence of reversible microscopic dynamics, such as that of hard spheres. In fact, Loschmidt correctly argued that if  $\mathcal{H}$  decreases in a time interval for a given choice of initial conditions, there is another



**Fig. 1.10** Schematic representation of TRI dynamics in phase space. The points of the continuous line represent the positions  $\mathbf{q}$ , the arrows represent the momenta  $\mathbf{p}$ ;  $\Gamma = (\mathbf{q}, \mathbf{p})$ . Time reversibility holds if the forward time trajectory with initial condition  $\Gamma'$  traces backward with opposite momenta the set of  $\mathbf{q}$ 's traced in the same time by the trajectory starting at  $\Gamma$ . The trajectories appear to overlap, but in phase space they do not, because they have different momenta

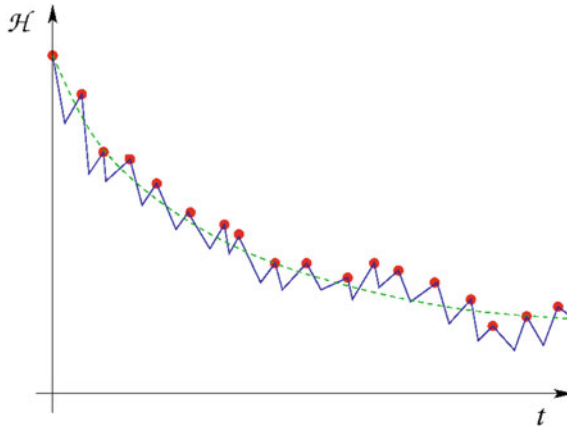
initial condition that leads  $\mathcal{H}$  to increase in the same time interval: *it suffices* to reverse the velocities of the second path.

The  $\mathcal{H}$ -theorem was declared impossible also on the grounds of the Poincaré recurrence theorem, that had been proven a shortly earlier. This objection, known as the *recurrence objection*, or *Zermelo paradox*, notices that given any finite tolerance, the phase space trajectories of mechanical systems with bounded phase space takes a finite time  $T_R$  to return to their initial condition within that tolerance. Therefore,  $\mathcal{H}$  may initially decrease but, being a continuous function of phase, sooner or later it will return as close as one wishes to its initial value, hence will sooner or later violate the monotonic behaviour implied by the Boltzmann equation.

Indeed, both objections are mathematically well motivated, but they do not consider a fact, which Boltzmann himself pointed out: a gas is not any kind of dynamical system, it is one with very many particles! While obvious, this fact has far-reaching consequences, that are not immediately clear in abstract mathematical terms, but that become evident when concrete numbers are given. For instance, the molecules of air at usual temperature and pressure in a volume of  $1 \text{ cm}^3$  have a typical  $T_R$  of order  $O(10^{10^{19}})$  years! It is also obvious that the difficulty in aiming the velocities of colliding spheres, so that they move from a disordered distribution to an ordered one increases exceedingly rapidly with the number  $N$  of balls. Furthermore, such an effort is totally pointless, because, even if successful for some time, after balls have grouped in a region of a billiard table, continuing their motion they move again apart from each other.

How can the formal correctness of Loschmidt and Zermelo's reasonings be reconciled with the most convincing and verifiable Boltzmann's theory?

Take  $\mathcal{H}$  for a system made of a number  $N$  of finite size balls in a given container with straight reflecting walls. In order to compute  $\mathcal{H}$ , the density  $\rho$  of the balls can be computed subdividing the container in volumes that are not too small, lest the quantity  $\rho$  turns nonsense, neither too large, corresponding to inaccurate resolution. Clearly  $N$  ought to be very large for both conditions to be met. If  $N$  is not sufficiently large, the time evolution of  $\mathcal{H}$  looks something like Fig.1.11. Therefore, given a system made of a macroscopic number of particles, neither the deviations from the monotonic decreasing curve, nor the increase due to recurrence will ever be observed, and Boltzmann theory is vindicated.



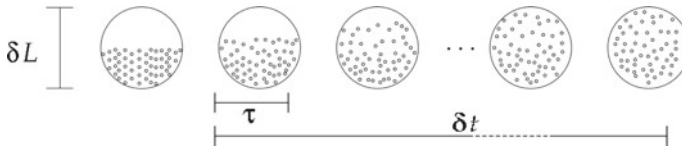
**Fig. 1.11** Evolution of Boltzmann  $\mathcal{H}$ -functional for a finite number  $N$  of particles, as obtained from their number density  $\rho$ . It only approximately follows the monotonic decrease that it would enjoy, were  $\rho$  be replaced by the solution  $f$  of the corresponding Boltzmann equation. Moreover, on a much longer scale  $O(T_R)$  not represented here,  $\mathcal{H}$  climbs up again to higher values, because of Poincaré’s recurrences. Larger  $N$  implies smaller fluctuations about the smooth curve, and longer times  $T_R$

### 1.7 Local Thermodynamic Equilibrium

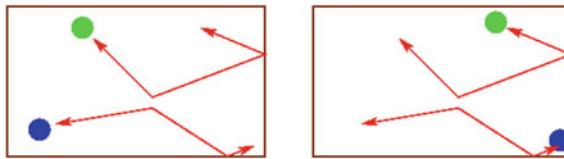
The transition from the discrete atomic description, to the continuum macroscopic description is strictly mathematically performed only within the kinetic theory of gases [29], as described in the previous Section. Nevertheless, the same ideas can be convincingly applied much more generally [30]. The main idea remains the same: one needs a wide separation of three length and time scales, called microscopic, mesoscopic and macroscopic scales. This, in turn, requires the object of interest to be made of a very large number,  $N \gg 1$ , of interacting atoms or molecules. Denoting by  $\ell$  and  $\tau$  the characteristic microscopic length and time, and by  $\delta L$  and  $\delta t$  the mesoscopic ones, it is meant that  $\delta L^3$  is sufficiently large that it may contain a small thermodynamic system, i.e. a sufficiently large number of particles that it make sense to assign to it properties such as pressure  $P$ , temperature  $T$  and density  $\rho$ . It is further meant that  $\delta t$  is long enough that thanks to interaction within  $\delta L^3$ , a homogeneous state is reached in  $\delta L^3$ . Denoting by  $L$  and  $t$  the macroscopic characteristic scales, i.e. the scales at which measurements take place, one further requires:

$$\ell \ll \delta L \ll L, \quad \tau \ll \delta t \ll t \tag{1.196}$$

In other words, Local Thermodynamic Equilibrium (LTE) is established, making a thermodynamic description feasible, if mesoscopic cells appear like points and reach equilibrium in a time  $\delta t$  that appear infinitesimal on the scale of measurements. This condition also amounts, from the macroscopic point of view, to a fast decays



**Fig. 1.12** LTE condition: the macroscopic system of interest is subdivided in cells, the circular volumes, are large compared to the microscopic space scales, so they may contain many particles, but are small compared to the macroscopic scales, the scales at which measurement takes place. LTE is established if the state of these cells becomes homogeneous in a time  $\delta t$  large compared to the microscopic times, but small compared to the macroscopic characteristic times. The condition  $\delta L \gg \ell$  ensures that many particles are contained in a volume  $\delta L^3$ , which experience many collisions in a short time. Collisions lead to a homogeneous state within the cell, which is then well represented by a single point, and also help particles in the bulk of a cell to remain inside that cell



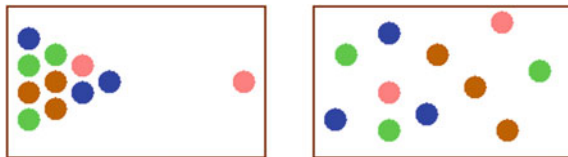
**Fig. 1.13** Two snapshots of a billiard table with only two balls. It is not possible to distribute the mass uniformly on the table; equivalently, it is not possible to identify the direction of time: which snapshot was taken earlier? They are equally plausible

of correlations both in space and in time, with cells surface effects negligible with respect to their bulk. Then, each  $\delta L^3$  volume may be considered as a small isolated equilibrium system. In the case of kinetic theory of gases,  $\ell$  is the mean free path and  $\tau$  the mean free time, cf. Fig. 1.12.

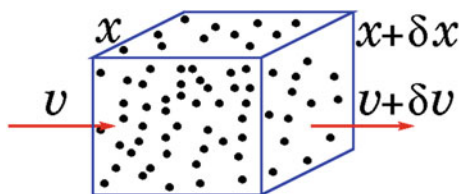
As is well known, a homogeneous probability distribution is achieved in the phase space of hard spheres, but that does not mean that a homogeneous mass distribution is automatically achieved in real space. Indeed, even a billiard with only two balls has the uniform probability distribution in phase space as its invariant probability density; but it cannot have a uniform mass distribution in real space. For that, large  $N$  is required as illustrated in Figs. 1.13 and 1.14.

This example shows that relaxation of mass distribution to the homogeneous state, which is an irreversible process required for LTE to be defined, is not possible unless the cells contain many particles, and the particles interact. In fact, in the case of two balls, while probability irreversibly relaxes to the uniform phase space distribution, no direction of time can be perceived in the motion of balls Fig. 1.13. At the same time, that balls are many does not guarantee that they converge to a (stable) uniform mass distribution; that requires collisions. In case of pointlike particles, an initial inhomogeneous mass distribution may be preserved in time.

This way, we have further demonstrated that the identification of probability in phase space and mass in real space is a delicate point of statistical mechanics. The first is an abstract notion referring to an hypothetical ensemble of identical objects;



**Fig. 1.14** Two snapshots of a billiard table with 11 balls. The notion of uniform mass distribution begins to make sense; it improves if the balls are smaller and a larger number. In case the motion is spontaneous and is subject to no friction, the direction of time is revealed by the transition from ordered to disordered (uniform) mass distribution. This is achieved for practically whatever initial impulse is given to a single ball. The reverse motion is not impossible, but exceedingly hard to achieve (all balls should be extremely precisely aimed). Moreover, the ordered distribution would turn into disorder, immediately after it has been created



**Fig. 1.15** Given a mesoscopic cube containing a very large number of very small particles, the granularity of matter can be neglected, and local balances of conserved quantities, such as mass, can be performed treating such quantities as a continuum, continuously flowing through the surface of the cube. Among other ingredients, this requires particles to interact, so that their motion is randomized and the in- and out-fluxes are negligible compared to the bulk

the second is a material measurable property of a single concrete object. Relaxation to LTE is not mere convergence to an invariant probability distribution.

When LTE holds, the *granularity* of the microscopic structure becomes irrelevant, on the scale of observation, analogously to the observation of a white cloud in the sky, that is made of droplets not seen from the ground. Matter can then be treated as a continuum, with continuously varying properties, and local balances of the quantities of interest can be performed, as we have done in Sect. 1.2, cf. Fig. 1.15.

The corresponding macroscopic description includes linear equations, like Fick’s law for the density  $n$  of tracer diffusion with diffusion coefficient  $D$ , or Ohm’s law for electric current  $J_e$  under an electric field  $E$  with conductivity  $\kappa$ :

$$J_n(x, t) = -D \frac{\partial n}{\partial x}(x, t), \quad J_e(x, t) = \kappa E \tag{1.197}$$

This currents, in turn, imply entropy sources of the form:

$$\sigma_n(x, t) = \frac{D}{n(x, t)} \left[ \frac{\partial n}{\partial x}(x, t) \right]^2, \quad \sigma_e(x, t) = \frac{J_e E}{k_B T} \tag{1.198}$$

Nonlinear generalizations also exist, but they still require LTE for the hydrodynamic/thermodynamic fields to exist.

For a concrete example, consider a piece of copper at 273 K [24]. One has:

$$\tau = 2.7 \cdot 10^{-14} \text{ s}$$

for the characteristic time with which electrons collisions with the atoms of the conductor lattice, and

$$E_F = \frac{1}{2} m_* v_F^2 = 7 \text{ eV}$$

which is the Fermi Level Energy of electrons of effective mass  $m_* = 1.3m_e$ ,<sup>18</sup> where

$$m_e = 0.511 \frac{M_e V}{c^2} = 9.1095 \cdot 10^{-31} \text{ kg}; \quad c = 2.998 \times 10^8 \text{ m/s} \Rightarrow$$

Then, the typical speed of the electron takes the value:

$$v_F = \sqrt{\frac{2E_F}{m_*}} = 13.763 \cdot \frac{10^8 \text{ m}}{10^3 \text{ s}} = 1.376 \times 10^6 \text{ m/s}$$

eventually leading to

$$\ell = v_F \tau = 1.376 \cdot 10^6 \frac{\text{m}}{\text{s}} \cdot 10^{-14} 2.7 \text{ s} = 3.716 \times 10^{-8} \text{ m}$$

Consequently, a number  $O(10^3)$  collisions take place in a cube of side  $\delta L = 3.716 \cdot 10^{-7} \text{ m}$ , every time  $\tau$ . Thus, taking  $\delta t \sim 10^2 \tau = 2.7 \cdot 10^{-12} \text{ s}$  guarantees relaxation to a homogeneous state. Then, because our observations may concern linear sizes of the order of millimeters and take times of the order of a second, LTE can be safely assumed.

Of course, LTE can be violated. For instance, in  $10^{-15} \text{ s}$  laser pulses; in nanometric devices; in cosmic rays hitting the screens of space ships having characteristic lengths larger than ship itself; in complex materials, such as proteins characterized by many more scales than 3, etc. In all these cases approaches that go beyond thermodynamics are necessary, although thermodynamic relations seem to apply quite more widely than expected.

---

<sup>18</sup>In periodic potential electrons are accelerated by electric fields as if they had a different mass  $m_*$ , that may be large or smaller than their mass  $m_e$ , and may even be negative.



## 1.8 Linear Response

The next natural question is what happens if an equilibrium state is perturbed by some external action. The simplest case consists of a particle system with Hamiltonian  $\mathcal{H}_0(\mathbf{P}, \mathbf{Q})$ , which is given some extra energy  $\lambda A(\mathbf{P}, \mathbf{Q})$  say, with  $\lambda \in \mathbb{R}$ , so that a new Hamiltonian  $\mathcal{H}(\mathbf{P}, \mathbf{Q}) = \mathcal{H}_0(\mathbf{P}, \mathbf{Q}) + \lambda A(\mathbf{P}, \mathbf{Q})$  is produced. If both the initial and the final states correspond to equilibria at inverse temperature  $\beta = 1/kT$ , the canonical ensembles:

$$f_0(\mathbf{P}, \mathbf{Q}) = \frac{\exp(-\beta\mathcal{H}_0)}{\int d\mathbf{P}d\mathbf{Q} \exp(-\beta\mathcal{H}_0)}, \quad f(\mathbf{P}, \mathbf{Q}) = \frac{\exp(-\beta\mathcal{H})}{\int d\mathbf{P}d\mathbf{Q} \exp(-\beta\mathcal{H})} \quad (1.199)$$

describe the statistics of the microscopic phases. Provided  $\lambda$  is small, the first order approximation in  $\lambda$  constitutes a good approximation of  $f$ :

$$f(\mathbf{P}, \mathbf{Q}) \simeq \frac{\exp(-\beta\mathcal{H}_0)}{\int d\mathbf{P}d\mathbf{Q} \exp(-\beta\mathcal{H}_0)} \frac{1 - \lambda\beta A(\mathbf{P}, \mathbf{Q})}{1 - \lambda\beta \langle A(\mathbf{P}, \mathbf{Q}) \rangle_0} \quad (1.200)$$

$$= f_0(\mathbf{P}, \mathbf{Q}) (1 - \lambda\beta [A(\mathbf{P}, \mathbf{Q}) - \langle A(\mathbf{P}, \mathbf{Q}) \rangle_0]) \quad (1.201)$$

where  $\langle \cdot \rangle_0$  means average with respect to the unperturbed ensemble  $f_0$ . This becomes the long time response of the system to the perturbation with small  $\lambda$ , if the state actually evolves from the equilibrium characterized by  $f_0$  to the equilibrium characterized by  $f$ . In the above derivation, there is no proof that  $f_0$  converges to  $f$ ; that depends on: **a**) very many details of the microscopic dynamics, that are mathematically hard to control, combined with **b**) the set of observables of interest, because convergence of observables is the only sense in which convergence of phase space probabilities can be understood.<sup>19</sup> However, when convergence does take place, one may use Eq. (1.201) to compute the variation of a generic observable  $\mathcal{O}$ , due to the small perturbation. The result is:

$$\langle \Delta\mathcal{O} \rangle_0 = \int d\mathbf{P}d\mathbf{Q} \mathcal{O}(\mathbf{P}, \mathbf{Q}) [f(\mathbf{P}, \mathbf{Q}) - f_0(\mathbf{P}, \mathbf{Q})] \quad (1.202)$$

$$\simeq -\lambda\beta (\langle \mathcal{O}A \rangle_0 - \langle \mathcal{O} \rangle_0 \langle A \rangle_0) \quad (1.203)$$

which means that the response of the variable  $\mathcal{O}$  is determined by the equilibrium correlation of  $\mathcal{O}$  and  $A$ . In the case that  $\mathcal{O} = A = \mathcal{H}_0$ , whose equilibrium average can be interpreted as the unperturbed internal energy, one obtains:

---

<sup>19</sup>To earlier observations, we add that the rule according to which phase points drag probability around makes it impossible, except in uninteresting situations, for the probability to come to rest in phase space.

$$\frac{\langle \Delta \mathcal{H}_0 \rangle_0}{\lambda \beta} \simeq -(\langle \mathcal{H}_0^2 \rangle_0 - \langle \mathcal{H}_0 \rangle_0^2) = -k_B T^2 C_V \quad \text{with } \langle \mathcal{H}_0^n \rangle_0 = \int \mathcal{H}_0(\mathbf{P}, \mathbf{Q})^n f_0(\mathbf{P}, \mathbf{Q}) d\mathbf{P} d\mathbf{Q} \quad (1.204)$$

where  $C_V$  is the heat capacity at constant volume. In other words, the response to energy perturbations, which defines the heat capacity, is linked to the equilibrium energy fluctuations. The heat capacity at constant volume is indeed the derivative of the internal energy with respect to the temperature, which can be implicitly obtained in the  $\lambda \rightarrow 0$  limit, considering that  $\beta = 1/k_B T$ . That we are dealing with an isochoric process is implicit in the fact that no other energy sources are considered. At constant pressure  $P$ , for instance, one would have to add a  $PdV$  contribution to the heat needed to change the temperature of the object of interest.

In the case of time dependent perturbations of the form  $-\mathcal{F}(t)A(\Gamma)$ :

$$H(\Gamma, t) = H_0(\Gamma) - \mathcal{F}(t)A(\Gamma) \quad (1.205)$$

where  $\mathcal{F}(t)$  is small, one may define the unperturbed and the perturbed evolution operators as:

$$i\mathcal{L}_0 f = \{f, H_0\}, \quad i\mathcal{L}_{\text{ext}}(t) f = -\mathcal{F}(t) \{f, A\} \quad (1.206)$$

where  $\{ \cdot \}$  are the Poisson brackets. If  $f_0$  is the unperturbed equilibrium, one has  $i\mathcal{L}_0 f_0 = 0$ , and the solution of the Liouville equation

$$\frac{\partial f}{\partial t} = -i(\mathcal{L}_0 + \mathcal{L}_{\text{ext}}(t)) f \quad (1.207)$$

can be expressed by [16]:

$$f_t(\Gamma) = e^{it\mathcal{L}_0} f_0(\Gamma) - i \int_0^t dt' e^{-i(t-t')\mathcal{L}_0} \mathcal{L}_{\text{ext}}(t') f_{t'}(\Gamma) \quad (1.208)$$

$$= f_0(\Gamma) - i \int_0^t dt' e^{-i(t-t')\mathcal{L}_0} \mathcal{L}_{\text{ext}}(t') f_0(\Gamma) + \text{higher order in } \mathcal{L}_{\text{ext}} \quad (1.209)$$

If the deviations from the unperturbed system are considered small, the higher orders in  $\mathcal{L}_{\text{ext}}$  can be omitted and the variation in time of the phase space average of  $\mathcal{O}$  is given by:

$$\langle \mathcal{O} \rangle_t - \langle \mathcal{O} \rangle_0 \simeq \int d\Gamma \mathcal{O}(\Gamma) \int_0^t dt' e^{-i(t-t')\mathcal{L}_0} \mathcal{F}(t') \{f_0, A\} \quad (1.210)$$

where

$$\{f_0, A\} = \{H_0, A\} \frac{\partial f_0}{\partial H_0} = \beta f_0 \frac{dA}{dt} \quad (1.211)$$

Eventually, one obtains:

$$\langle \mathcal{O} \rangle_t - \langle \mathcal{O} \rangle_0 \simeq \int_0^t dt' R(t-t') \mathcal{F}(t') \quad (1.212)$$

where  $R(t)$  is the response function:

$$R(t) = \beta \langle \dot{A}(\mathcal{O} \circ S^t) \rangle_0 = \beta \int d\Gamma f_0(\Gamma) \frac{dA}{dt}(\Gamma) e^{it\mathcal{F}_0} \mathcal{O}(\Gamma) \quad (1.213)$$

Once again, the macroscopic nonequilibrium behaviour of a given system has been related solely to the correlations of microscopic fluctuating quantities, computed with respect to the relevant equilibrium ensemble.

Equation (1.212) suggests that even the linear response is in general affected by memory effects. From this point of view, the Markovian behaviour seems to be either very special or a crude approximation, implying, for instance, that all nonequilibrium fluids have a viscoelastic behaviour. In practice, however, in normal fluids and normal conditions the memory terms decay rapidly, so that the Markovian approximation is by and large justified. The viscoelastic behaviour is indeed noticeable only in complex fluids or under extreme conditions, i.e. exceedingly far from equilibrium.

For example, perturb  $\mathcal{H}_0$  with a spatially uniform and constant force  $h$  along the  $x$  direction. For small  $h$ , the average velocity moderately varies in time, and the overall current  $\mathcal{J}(\mathbf{P}, \mathbf{Q}) = \frac{1}{m} \sum_j p_j^x$ , can be considered to linear order:

$$\langle \mathcal{J} \rangle_t - \langle \mathcal{J} \rangle_0 = \beta \frac{h}{m^2} \int_0^t dt' \sum_{j,k} \langle p_j^x (p_k^x \circ S^{t-t'}) \rangle_0$$

Assuming that momenta of different particles are uncorrelated at equilibrium, and recalling that the mobility  $\mu$  is defined by

$$\lim_{t \rightarrow \infty} \langle \Delta \mathcal{J} \rangle_t = \lim_{t \rightarrow \infty} \langle \mathcal{J} \rangle_t - \langle \mathcal{J} \rangle_0 = \mu h$$

one eventually obtains

$$\mu = \frac{\beta}{m^2} \int_0^\infty dt' \sum_j \langle p_j^x(0) p_j^x(t') \rangle_0 \quad (1.214)$$

which is the Green-Kubo relation for diffusion, related to the diffusion coefficient  $D$  by the Einstein relation  $D = \mu/\beta$ . Then, writing Eq. (1.214) as

$$\langle \Delta \mathcal{J} \rangle_t = \int_{t_0}^t dt' R(t-t') \mathcal{F}(t') \quad (1.215)$$

one observes that the response function takes the form:

$$R(t) = \beta \langle \dot{A}(\mathcal{J} \circ S^t) \rangle = -\beta \langle A(\dot{\mathcal{J}} \circ S^t) \rangle_0 \quad (1.216)$$

These relations, as previously noted, are universally confirmed, which is an indirect way of proving the validity of the linear response theory. Indeed, linear response theory is extremely successful and can be analogously derived for stochastic process, just replacing the phase space with the state space and the Liouville equation with e.g. the Fokker-Planck equation. Compare the results of this section with those of Sect. refFDRsect and, in particular with Eqs. (1.151), (1.160).

### 1.8.1 Modern Developments of Linear Response

The above formalism can be extended to perturbations of nonequilibrium steady states. For instance, to steady states whose dynamics is dissipative, hence not Hamiltonian, as in the important case of viscous hydrodynamics [31]. Recently, it has been shown that the approach we have outlined does indeed apply, if the steady state is represented by a regular probability density, as commonly happens in the presence of noise, cf. Refs. [31, 32].

Differently, the invariant phase space probability distribution of a dissipative system  $\mu$ , say, is typically singular and supported on a fractal attractor. Consequently, it is not obvious anymore that the statistical features induced by a perturbation can be related to the unperturbed statistics. The reason is that even very small perturbations may lead to microscopic states whose probability vanishes in the unperturbed state  $\mu$ . In such a case, the information contained in  $\mu$  is irrelevant.

Indeed, Ruelle [33] showed that in certain cases<sup>20</sup> a perturbation  $\delta\Gamma$  about a microstate  $\Gamma$  and its evolution  $S^t\delta\Gamma$  can be decomposed in two parts,  $(S^t\delta\Gamma)_\parallel$  and  $(S^t\delta\Gamma)_\perp$ , respectively perpendicular and parallel to the fibres of the attractor:

$$S^t\delta\Gamma = (S^t\delta\Gamma)_\parallel + (S^t\delta\Gamma)_\perp$$

The first addend can be related to the dynamics on the attractor, while the second may not.

Later, it has been pointed out [34] that this difficulty should not concern the systems of many interacting particles which are of statistical mechanics interest. In those cases, rather than the full phase space, one considers the much lower dimensional projections, afforded by a few physically relevant observables. Hence, one is typically

---

<sup>20</sup>Concerning certain smooth, uniformly hyperbolic dynamical systems.

interested in the marginals of singular phase space measures, on spaces of sufficiently lower dimension, which are usually regular [35, 36]. These facts can be briefly recalled as follows. Ruelle showed that the effect of a perturbation  $\delta F(t) = \delta F_{\parallel}(t) + \delta F_{\perp}(t)$  on the response of a generic (smooth enough) observable  $\mathcal{O}$  is given by:

$$\langle \mathcal{O} \rangle_t - \langle \mathcal{O} \rangle_0 = \int_0^t R_{\parallel}^{(\mathcal{O})}(t - \tau) \delta F_{\parallel}(\tau) d\tau + \int_0^t R_{\perp}^{(\mathcal{O})}(t - \tau) \delta F_{\perp}(\tau) d\tau \quad (1.217)$$

where the subscript 0 denotes averaging with respect to  $\mu$ ,  $R_{\parallel}^{(\mathcal{O})}$  may be expressed in terms of correlation functions evaluated with respect to  $\mu$ , while  $R_{\perp}^{(\mathcal{O})}$  depends on the dynamics along the stable manifold, hence it may not.

Let us adopt the point of view of Ref. [34]. For a  $d$ -dimensional dissipative dynamical system consider, for simplicity, an impulsive perturbation  $\Gamma \rightarrow \Gamma + \delta\Gamma$ , such that all components of  $\delta\Gamma$  vanish except one, denoted by  $\delta\Gamma_i$ . The probability distribution  $\mu$  is correspondingly shifted by  $\delta\Gamma$ , and turns into a non-invariant distribution  $\mu_0$ , whose evolution  $\mu_t$  tends to  $\mu$  in the  $t \rightarrow \infty$  limit. For every measurable set  $E \subset \mathcal{M}$ ,  $\mu_0(E)$  is given by  $\mu(E - \delta\Gamma)$ ,<sup>21</sup> and  $\mu_t(E)$  is computed as explained in Sec. 1.6. Taking  $\mathcal{O}(\Gamma) = \Gamma_i$ , one obtains:

$$\langle \Gamma_i \rangle_t - \langle \Gamma_i \rangle_0 = \int \Gamma_i d\mu_t(\Gamma) - \int \Gamma_i d\mu(\Gamma) \quad (1.218)$$

Approximate the singular  $\mu$  by means of piecewise constant distributions, introducing an  $\epsilon$ -partition made of a finite set of  $d$ -dimensional hypercubes  $\Lambda_k(\epsilon)$  of side  $\epsilon$  and centers  $\Gamma_k$ . We define an  $\epsilon$ -approximation of  $\mu$  and of  $\mu_t$  in terms of the probabilities  $P_k(\epsilon)$  and  $P_{t,k}(\epsilon; \delta\Gamma)$  of the hypercubes  $\Lambda_k(\epsilon)$ :

$$P_k(\epsilon) = \int_{\Lambda_k(\epsilon)} d\mu(\Gamma), \quad P_{t,k}(\epsilon) = \int_{\Lambda_k(\epsilon)} d\mu_t(\Gamma). \quad (1.219)$$

This yields the coarse grained invariant density  $\rho(\Gamma; \epsilon)$ :

$$\rho(\Gamma; \epsilon) = \sum_k \rho_k(\Gamma; \epsilon), \quad \text{with } \rho_k(\Gamma; \epsilon) = \begin{cases} P_k(\epsilon)/\epsilon^d & \text{if } x \in \Lambda_k(\epsilon) \\ 0 & \text{else} \end{cases} \quad (1.220)$$

If  $Z_i$  is the number of one-dimensional bins of form  $\left[ \Gamma_i^{(q)} - \epsilon/2, \Gamma_i^{(q)} + \epsilon/2 \right)$ ,  $q \in \{1, 2, \dots, Z_i\}$ , in the  $i$ -th direction, marginalizing the approximate distribution yields the following quantities:

<sup>21</sup>The set  $E - \delta\Gamma$  is defined by  $\{\Gamma \in \mathcal{M} : \Gamma + \delta\Gamma \in E\}$ .

$$p_i^{(q)}(\epsilon) = \int_{\Gamma_i^{(q)} - \frac{\epsilon}{2}}^{\Gamma_i^{(q)} + \frac{\epsilon}{2}} \left\{ \int \rho(\Gamma; \epsilon) \prod_{j \neq i} d\Gamma_j \right\} d\Gamma_i \quad (1.221)$$

Each of them is the invariant probability that the coordinate  $\Gamma_i$  of  $\Gamma$  lie in one of the  $Z_i$  bins. Similarly, one gets the marginal of the evolving approximate probability  $p_{t,i}^{(q)}(\epsilon)$ . In both cases, dividing by  $\epsilon$ , one obtains the coarse grained marginal probability densities  $\rho_i^{(q)}(\epsilon)$  and  $\rho_{t,i}^{(q)}(\epsilon)$ , as well as the  $\epsilon$ -approximate response function:

$$B_i^{(q)}(\Gamma_i, \delta\Gamma, t, \epsilon) = \frac{1}{\epsilon} \left[ p_{t,i}^{(q)}(\epsilon) - p_i^{(q)}(\epsilon) \right] = \rho_{t,i}^{(q)}(\epsilon) - \rho_i^{(q)}(\epsilon) \quad (1.222)$$

Reference [34] shows that the right hand side of Eq. (1.222) tends to a regular function of  $\Gamma_i$  under the  $Z_i \rightarrow \infty$ ,  $\epsilon \rightarrow 0$  limits. Consequently,  $B_i^{(q)}(\Gamma_i, \delta\Gamma, t, \epsilon)$  yields an expression similar to that of standard response theory, in the sense that it depends solely on the unperturbed state, although that is supported on a fractal set. There are exceptions to this conclusion, most notably those discussed by Ruelle. But for systems of many interacting particles this is the expected result. The idea is that the projection procedure makes unnecessary the explicit calculation of  $R_{\perp}^{(O)}$  in Eq. (1.217). This does not mean that  $R_{\perp}^{(O)}$  is necessarily negligible [37]. However, apart from peculiar situations, it does not need to be explicitly computed and the response may be referred only to the unperturbed dynamics, as in the standard theory.

## 1.8.2 Linear Response in Magnetic Field

One of the major results of linear nonequilibrium thermodynamics are the Onsager reciprocal relations. Given a system subjected to a set of driving forces  $X_j$ , each of which separately induces a current  $J_j$ ,  $j = 1, \dots, n$ , the linear regime is characterized by the following relation:

$$J_i = \sum_{j=1}^n L_{ij} X_j, \quad X_i = \sum_{j=1}^n R_{ij} J_j \quad (1.223)$$

Under the hypothesis that LTE holds, and that the microscopic dynamics are TRI, Onsager proved that that matrix of transport coefficients ( $L_{ij}$ ) is symmetric:

$$L_{ij} = L_{ji}, \quad i, j = 1, \dots, n \quad (1.224)$$

The set of equations (1.224) is called Onsager Reciprocal Relations, and they constitute the theoretical expression of phenomena that had been previously observed, such as the Soret-Dufour effect, concerning a mixture in which neither temperature

nor concentration are uniform. In this case one realizes that the flow of heat gets a contribution from the thermodynamic force due to the concentration gradient, with proportionality constant  $L_{qm}$ , and the diffusion of mass gets a contribution from the thermodynamic force associated with the temperature gradient, with proportionality constant  $L_{mq}$ . It turns out that  $L_{qm} = L_{mq}$ . Similarly, the thermo-electric effect, or Peltier-Seebeck-Thomson effect, couple heat flows with electric currents.

Onsager relations introduce constraints on the conversion of heat to work, like other thermodynamic relations, therefore their violation may in principle allow non-dissipative thermodynamic currents. At the same time, it was argued, and universally accepted, that the Onsager reciprocal relations do not hold in systems subjected to a magnetic field, or a rotating reference frame, because their dynamics are not TRI. Casimir then argued the Onsager relations should be replaced by the following [38]:

$$L_{ij}(\mathbf{B}) = L_{ji}(-\mathbf{B}), \quad i, j = 1, \dots, n \tag{1.225}$$

if  $\mathbf{B}$  is the magnetic field in which the system of interest is immersed. While this relation is conceptually satisfactory, it defeats the purpose of obtaining the transport coefficient for a given system: two systems, one in the magnetic field  $\mathbf{B}$  and the other in the magnetic field  $-\mathbf{B}$ , must be considered. Then, the  $L_{ij}$  coefficient of one system cannot be inferred from the  $L_{ji}$  coefficient of that same system: a second experiment, with a different system, must be performed.

However, non-dissipative currents have not been observed, so far, hence the question whether magnetic fields break the Onsager symmetry has remained open. Recently, this question has received an answer, in favour of the validity of Onsager relations even in the presence of magnetic fields, or rotating frames, because in reality they do not break all possible time reversal symmetries [39]. This has direct application on the correlation functions from which the transport coefficients can be computed, via the Green-Kubo relations. In Ref. [39] it has been shown that for systems whose particles interactions only depend on their relative positions, there are at least 7 more time reversal operators that correspond to TRI. For instance, the following operation:

$$i(x, y, z, p_x, p_y, p_z) = (x, -y, z, -p_x, p_y, -p_z) \tag{1.226}$$

had been used in the case of shearing fluids. It preserves the form of the equations of motion in presence of a magnetic field, if applied together with time inversion  $t \rightarrow -t$ :

$$\begin{array}{ll}
\dot{x}_i = \frac{p_{ix}}{m_i} + \omega_i y_i & \longmapsto \quad -\dot{x}_i = -\frac{p_{ix}}{m_i} - \omega_i y_i \\
\dot{y}_i = \frac{p_{iy}}{m_i} - \omega_i x_i & \longmapsto \quad \dot{y}_i = \frac{p_{iy}}{m_i} - \omega_i x_i \\
\dot{z}_i = \frac{p_{iz}}{m_i} & \longmapsto \quad -\dot{z}_i = -\frac{p_{iz}}{m_i} \\
\dot{p}_{ix} = F_{ix} + \omega_i (p_{iy} - m_i \omega_i x_i) & \longmapsto \quad \dot{p}_{ix} = F_{ix} + \omega_i (p_{iy} - m_i \omega_i x_i) \\
\dot{p}_{iy} = F_{iy} - \omega_i (p_{ix} + m_i \omega_i y_i) & \longmapsto \quad -\dot{p}_{iy} = -F_{iy} + \omega_i (p_{ix} + m_i \omega_i y_i) \\
\dot{p}_{iz} = F_{iz} & \longmapsto \quad \dot{p}_{iz} = F_{iz}
\end{array}$$

The rule is that one may change or not change the sign of one coordinate, as long as the opposite is done with the corresponding momentum. Because there are 3 coordinates and 2 signs, the number of allowed transformations of this kind is  $2^3 = 8$ ; four of them preserve the equations of motion, in the presence of a magnetic field, but they do not include the standard one

$$(x, y, z, p_x, p_y, p_z) \longmapsto (x, y, z, -p_x, -p_y, -p_z) \quad (1.227)$$

Interestingly, an electric field also breaks four of the symmetries and preserves the remaining four, but these include the standard one (1.227).

For any time reversal operation, the calculation of the equilibrium time correlation functions proceeds as follows: take the equilibrium probability density in phase space, which obeys  $f(i\Gamma) = f(\Gamma)$ , and take two observables,  $\Phi$  and  $\Psi$  say, that obey  $\Phi(i\Gamma) = \eta_\Phi \Phi(\Gamma)$  and  $\Psi(i\Gamma) = \eta_\Psi \Psi(\Gamma)$ , where  $\eta_\mathcal{O} = \pm 1$  is the signature of  $\mathcal{O}$  under the time reversal operation  $i$ . Then, one can write:

$$\begin{aligned}
\langle \Phi(\Psi \circ S_{\mathbf{B}}^t) \rangle_{\mathbf{B}} &= \int dX \rho(X) \Phi(X) \Psi(S_{\mathbf{B}}^t X) = \int dY \rho(iY) \Phi(iY) \Psi(S_{\mathbf{B}}^t iY) \quad (1.228) \\
&= \eta_\Phi \int dY \rho(Y) \Phi(Y) \Psi(i S_{\mathbf{B}}^{-t} Y) = \eta_\Phi \eta_\Psi \int dY \rho(Y) \Phi(Y) \Psi(S_{\mathbf{B}}^{-t} Y) \\
&= \eta_\Phi \eta_\Psi \langle \Phi(0) (\Psi \circ S_{\mathbf{B}}^{-t}) \rangle_{\mathbf{B}} \quad (1.229)
\end{aligned}$$

where  $S_{\mathbf{B}}^t$  is the time evolution in presence of the magnetic field  $\mathbf{B}$ . This does not only show that Onsager reciprocal relations also hold in presence of a constant magnetic field, but adds predictive power to the theory. For instance, given a second admissible time reversal operation,  $\bar{i}$  say, let the corresponding signatures be  $\epsilon_\Psi$  and  $\epsilon_\Phi$ . Then, both relation

$$\langle \Phi(\Psi \circ S_{\mathbf{B}}^t) \rangle_{\mathbf{B}} = \eta_\Phi \eta_\Psi \langle \Phi(\Psi \circ S_{\mathbf{B}}^{-t}) \rangle_{\mathbf{B}} = \eta_\Phi \eta_\Psi \langle (\Phi \circ S_{\mathbf{B}}^t) \Psi \rangle_{\mathbf{B}} \quad (1.230)$$

$$\langle \Phi(\Psi \circ S_{\mathbf{B}}^t) \rangle_{\mathbf{B}} = \epsilon_\Phi \epsilon_\Psi \langle \Phi(\Psi \circ S_{\mathbf{B}}^{-t}) \rangle_{\mathbf{B}} = \epsilon_\Phi \epsilon_\Psi \langle (\Phi \circ S_{\mathbf{B}}^t) \Psi \rangle_{\mathbf{B}} \quad (1.231)$$

hold. Then either  $\eta_\Phi \eta_\Psi = \epsilon_\Phi \epsilon_\Psi$ , or



$$\langle \Phi (\Psi \circ S_{\mathbf{B}}^t) \rangle_{\mathbf{B}} = 0 \quad (1.232)$$

implying e.g. that certain transport coefficients vanish, i.e. that certain thermodynamic forces do not contribute to certain currents.

## 1.9 Beyond Linearity: Anomalies and Fluctuation Relations

We have learned that under the LTE condition there is a very successful theory, based on hydrodynamic and thermodynamic laws. In these conditions, the shape of the container of a fluid, for instance, does **not** matter: the boundaries do not affect the transport laws, they only appear as boundary conditions that must be imposed to solve the differential equations that represent the physical laws. In our daily life, it is definitely hard to break LTE and continuum mechanics that applies to transport processes, pattern formation, turbulence, etc. All that is based on the linear response theory; for instance transport of heat is proportional to temperature gradients via thermal conductivity; turbulence is described by the Navier-Stokes equations, in which the viscosity constitutes the proportionality between velocity gradients and transport of momentum in fluids. Nonlinear extensions of such transport laws, are envisaged under the LTE assumption that guarantees the existence of the hydrodynamic and thermodynamic fields; however, such nonlinear extensions do not come close to the wide applicability and success of linear relations.

Beyond LTE, the kinetic theory of gases remains applicable; one reason is that it holds when only two scales, rather than three, are sufficiently widely separated, e.g.  $\ell \ll L$ . Even in that case, walls only appear as boundary conditions. However, even the Boltzmann equation rests on unobvious conditions: the stosszahl-ansatz, which requires inter-particles interactions, and the very large number  $N$  of microscopic components.

Therefore, even the conditions that pertain to the kinetic theory of gases can be violated. For instance,  $N$  may be relatively small, or the particles may be highly confined, so that they interact more with the walls of their container than with each other. In this case, both inter-particle and particle-wall interactions determine the transport law, up to the extreme case in which particles only interact with their container, which then fully determines the transport law, often in very unstable manners [40, 41]. These are the cases in which fluctuations dominate, correlations persist in time and space, and transport is typically *anomalous*. This means that, unlike standard diffusion, transport is characterized by a nonlinear growth of the mean square displacement:

$$\langle [x(t) - x(0)]^2 \rangle \simeq t^\gamma, \quad \gamma \in [0, 2] \quad (1.233)$$

For  $\gamma < 1$ , one speaks of subdiffusion,  $\gamma = 1$  corresponds to normal diffusion and  $\gamma > 1$  is called super-diffusion.

The fields in which anomalous transport is realized are exploding; a brief list includes controlled drug delivery; photons in inhomogeneous media; running sand

piles; laser-cooled atoms; particles inside living cancer cells; particles passively advected by dynamical membranes; in the bulk-mediated diffusion on lipid bilayers; transport of heat in nanotubes, etc. Levy flights and walks are among the most common stochastic models of such systems; nonlinear 1D oscillators chains and polygonal billiards are among the deterministic models; exactly solvable models, meant to understand the basic mechanisms, are also available [41].

### 1.9.1 Transient and Steady State Fluctuation Relations

Arguably, the major results of nonequilibrium statistical physics of the past three decades concern an extensions of the equilibrium fluctuation theory, that originated from a paper by Evans, Cohen and Morriss dated 1993 [42]. That paper considered a TRI, but dissipative model of shearing fluids, known as SLLD, which is a special case of the following models:

$$\dot{\mathbf{q}}_i = \frac{\mathbf{p}_i}{m} + \mathbf{C}_i \mathbf{F}_e; \quad \dot{\mathbf{p}}_i = \mathbf{F}_i + \mathbf{D}_i \mathbf{F}_e - \alpha \mathbf{p}_i, \quad i = 1, \dots, N \quad (1.234)$$

Then, they proposed and tested the Fluctuation Relation (FR) that was later recognized as a generalization of Green-Kubo relations. Informally, this first FR can be written as:

$$\frac{\text{Prob.}(\overline{\Sigma}_\tau \approx -A)}{\text{Prob.}(\overline{\Sigma}_\tau \approx A)} \approx \exp[-A\tau] \quad (1.235)$$

where  $\overline{\Sigma}_\tau$  is the average energy dissipation rate in long time intervals of duration  $\tau$ , in a steady state. Given the reversibility of the dynamics, the dissipation may fluctuate and even take negative values; but the formula states that negative values occur with a probability that, compared to that of positive values, is exponentially small both in the size of the dissipation and of the observation times. This, analogously to Boltzmann's arguments on the  $H$  theorem, explains the second law of thermodynamics for these kinds of systems. Because  $\sigma$  is an extensive quantity, its values are proportional to the numbers of particles, hence very large in microscopic terms, like the observation times are also very large compared to microscopic times. The exponential of minus the product of these two quantities is thus practically zero, which means that in macroscopic experiments the entropy production can never be negative.

This relation, however, does not need to be related to thermodynamic phenomena. It represents a property of a reversible and dissipative particle system in a nonequilibrium steady state. In particular, it may apply to small systems, or to observations of short duration, or both. In that case, the fluctuations that are not observable in macroscopic systems may be observable. This is the case of nano-tech and bio-physical systems. It is also the case of macroscopic systems observed at a microscopic scale, such as the gravitational wave detectors. First derived within the deterministic molecular

dynamics framework, the FR was subsequently confirmed in the stochastic frameworks of Master and Langevin equations, but it is far from a generic property of physical systems. Below, we summarize the derivation given in Ref. [43]. As done in Section 1.6, for a dynamical system

$$\dot{\Gamma} = G(\Gamma), \quad \text{in the phase space } \mathcal{M} \quad (1.236)$$

we denote by  $S^t \Gamma$  the position in  $\mathcal{M}$  at time  $t$ , for a trajectory starting at  $\Gamma = (\mathbf{q}, \mathbf{p})$  at time 0. Moreover, let  $f^{(0)}$  be an initial probability distribution on  $\mathcal{M}$ , that is even with respect to the time reversal operation:  $f^{(0)}(i\Gamma) = f^{(0)}(\Gamma)$ . Then, the dissipation function  $\Omega^{f^{(0)}}$  defined by Eq. (1.171) is odd with respect to the time reversal operation,  $\Omega^{f^{(0)}}(i\Gamma) = -\Omega^{f^{(0)}}(\Gamma)$  as the energy dissipation should be. In fact,  $\Omega^{f^{(0)}}$  for standard models of nonequilibrium molecular dynamics equals the energy dissipation rate

$$\Sigma = FJ/k_b T \quad (1.237)$$

where  $F$  is a driving external force, and  $J$  the corresponding dissipative current. In the case of local equilibrium, this is the entropy production rate. The function  $\Omega^{f^{(0)}}$  is odd with respect to the time. Let us also assume that this dynamical system is TRI, in the sense of Eq. (1.195), although it may be dissipative, in the sense that the phase space volume variation rate is negative on average:

$$\lim_{t \rightarrow \infty} \langle \Lambda \rangle_t = \lim_{t \rightarrow \infty} \langle \text{div} G \rangle_t < 0 \quad (1.238)$$

where the average at time  $t$  is given by the probability distribution derived by evolving  $f^{(0)}$  for a time  $t$ .

Let us denote by  $A_\delta^+ = (A - \delta, A + \delta)$  and  $A_\delta^- = (-A - \delta, -A + \delta)$  two symmetric intervals of values that  $\Omega^{f^{(0)}}$  can take, and observe that TRI implies:

$$\{\Gamma : \overline{\Omega}_{0,\tau}^{(0)}(\Gamma) \in A_\delta^-\} = i S^\tau \{\Gamma : \overline{\Omega}_{0,\tau}^{(0)}(\Gamma) \in A_\delta^+\} \quad (1.239)$$

where the subscripts of the phase functions indicate integration in time as defined by Eq. (1.173). Consider the ratio of probabilities of opposite values of  $\Omega^{f^{(0)}}$ , computed with respect to  $f^{(0)}$ :

$$\frac{\mu^{(0)}(\overline{\Omega}_{0,\tau}^{(0)} \in A_\delta^+)}{\mu^{(0)}(\overline{\Omega}_{0,\tau}^{(0)} \in A_\delta^-)} = \frac{\int_{A_\delta^+} f^{(0)}(\Gamma) d\Gamma}{\int_{A_\delta^-} f^{(0)}(\Gamma) d\Gamma} \quad (1.240)$$

and introduce the coordinate transformation  $\Gamma = i S^\tau X$ , with jacobian

$$J_{0,\tau}(X) = \left| \frac{d\Gamma}{dX} \right| = \exp \{ \Lambda_{0,\tau}(X) \} \quad (1.241)$$

that leads to:

$$\int_{A_\delta^-} f^{(0)}(\Gamma) d\Gamma = \int_{A_\delta^+} f^{(0)}(iS^\tau X) e^{\Lambda_{0,\tau}(X)} dX =$$

$$\int_{A_\delta^+} f^{(0)}(X) e^{-\Omega_{0,\tau}^{(0)}(X)} dX = e^{-[A+\epsilon(A,\delta,\tau)]\tau} \int_{A_\delta^+} f^{(0)}(X) dX$$

One may thus write:

$$\frac{\mu^{(0)}(\overline{\Omega}_{0,\tau}^{(0)} \in A_\delta^+)}{\mu^{(0)}(\overline{\Omega}_{0,\tau}^{(0)} \in A_\delta^-)} = \exp\{\tau[A + \epsilon(A, \delta, \tau)]\}; \quad \epsilon(A, \delta, \tau) \leq \delta \quad (1.242)$$

where the function  $\epsilon$  is a correction term that can be made as small as one wants, reducing the width  $\delta$  of the observed values. This result is known as the Transient  $\Omega$ -FR. It is a remarkable identity that has been obtained under very minimal assumptions, i.e. the parity of the initial ensemble, which is verified by equilibrium states, and the TRI of the following nonequilibrium dynamics. Therefore, it is quite unbreakable and it holds for all observation times  $\tau$ , long or short they may be. It is called transient because it describes an ensemble of experiments starting in the same macroscopic state represented by  $f^{(0)}$ , typically an equilibrium state as e.g. in the Jarzynski equality, but with different initial microscopic state  $\Gamma$ . In turn, the Jarzynski equality is expressed by:

$$\langle e^{-\beta W} \rangle_A = e^{-\beta[F(B) - F(A)]} \quad (1.243)$$

where  $F(B) - F(A)$  is the free energy difference between two equilibrium states at inverse temperature  $\beta$ , characterized by two values of a parameter  $\lambda$ ,  $W$  is the work done to change  $\lambda$  from its value  $A$  to its value  $B$ , and the average is taken over the initial canonical ensemble with  $\lambda = A$ . The applicability of this equality still poses interesting questions, and it is object of current research.

The transient FR (1.242) has been verified, e.g. in experiments in which optical tweezers trap or drive colloidal particles [44]. Perhaps, it is conceptually most interesting because it closes circle with the fluctuation dissipation relation: the transient FR obtains information about the equilibrium state by performing nonequilibrium experiments, while the fluctuation dissipation relation obtains nonequilibrium properties such as the viscosity from equilibrium experiments.

Transient FRs are subtly but quite different from the original one of Ref. [42], that concerned steady states, despite their similar exponential aspect. Steady state FR, indeed, do not need to refer to ensembles of experiments: they express the fluctuations in time of the energy dissipation of a single object, in a stationary state. Not only the experiment is different, but also the statistics are different, since they refer to the steady state and not to the initial equilibrium state. The two are not matched

by taking long  $\tau$ , because the transient FR refers to the initial state, however long  $\tau$  might be. Nevertheless, under certain conditions that also guarantee convergence to the steady state, the transient FR turns into the steady state FR. The following derivation explains how this happens. Using the conservation of probability in phase space, that states

$$\mu^{(t)}(E) = \mu^{(0)}(S^{-t}E), \quad \text{for any measurable set } E \in \mathcal{M} \quad (1.244)$$

we may advance the initial distribution  $\mu^{(0)}$  up to time  $t$ , take the log of Eq. (1.242) and divide it by  $\tau$ , which yields:

$$\begin{aligned} \frac{1}{\tau} \ln \frac{\mu^{(t)}(\overline{\Omega}_{0,\tau}^{(0)} \in A_{\delta}^{+})}{\mu^{(t)}(\overline{\Omega}_{0,\tau}^{(0)} \in A_{\delta}^{-})} &= -\frac{1}{\tau} \ln \left\langle e^{-\Omega_{0,t}^{(0)}} \cdot e^{-\Omega_{t,t+\tau}^{(0)}} \cdot e^{-\Omega_{t+\tau,2t+\tau}^{(0)}} \right\rangle_{\overline{\Omega}_{t,t+\tau}^{(0)} \in A_{\delta}^{+}}^{(0)} \\ &= A + \epsilon(\delta, t, A, \tau) - \frac{1}{\tau} \ln \left\langle e^{-\Omega_{0,t}^{(0)}} \cdot e^{-\Omega_{t+\tau,2t+\tau}^{(0)}} \right\rangle_{\overline{\Omega}_{t,t+\tau}^{(0)} \in A_{\delta}^{+}}^{(0)} \end{aligned} \quad (1.245)$$

where  $\epsilon(\delta, t, A, \tau) \leq \delta$ , and  $\langle \cdot \rangle_{\overline{\Omega}_{t,t+\tau}^{(0)} \in A_{\delta}^{+}}^{(0)}$  denotes the average with respect to the initial distribution  $\mu^{(0)}$ , under the condition

$$\overline{\Omega}_{t,t+\tau}^{(0)}(\Gamma) \in A_{\delta}^{+} \quad (1.246)$$

This relation is exact, like Eq. (1.242) is. Taking  $t \rightarrow \infty$ , one should obtain the asymptotic expression for the steady state distribution  $\mu_{\infty}$ . Then, letting  $\tau \rightarrow \infty$  should eliminate the conditional average, producing the steady state FR. There is however a difficulty: the  $t \rightarrow \infty$  limit does not need to exist: the exponentials inside the conditional average could diverge, because they contain integrals from 0 to  $t$  of  $\overline{\Omega}_{t,t+\tau}^{(0)}$ . On the other hand, one cannot take the  $\tau \rightarrow \infty$  limit first, because infinitely long time averages may collapse on a single value; that value then gets probability 1, and all the others get probability 0, making nonsensical the FR expression. What makes sense is the steady state probability, if any, of the finite time averages of  $\overline{\Omega}_{t,t+\tau}^{(0)}$ , which can be positive; therefore,  $t$  has to grow first. When the  $\mu_{\infty}$  probabilities exist, one may ask what their ratio does, for longer and longer observation times  $\tau$ .

A common way to proceed, at this point, is to search for the conditions under which

$$M(A, \delta, \tau, t) = \ln \left\langle e^{-\Omega_{0,t}^{(0)}} \cdot e^{-\Omega_{t+\tau,2t+\tau}^{(0)}} \right\rangle_{\overline{\Omega}_{t,t+\tau}^{(0)} \in A_{\delta}^{+}}^{(0)} \quad (1.247)$$

does not diverge when  $t$  grows without bounds. Unfortunately, typically one obtains sufficient conditions, which may be too restrictive to be physically relevant. One example of that is the Anosov property, which can hardly be verified by systems of physical interest. The drawback of this approach is that one may miss the mechanism leading to the validity of the desired result, mingling it with unnecessary ingredients. A different path has been followed in Ref. [43], where necessary conditions have

instead been identified. In particular, for the validity of the standard steady state FR, one needs

$$\lim_{t \rightarrow \infty} M(A, \delta, \tau, t) \quad (1.248)$$

to exist, in which case we denote by  $\tilde{M}(A, \delta, \tau)$ , or at least to remain bounded. When that happens for any  $\delta > 0$ , the long  $\tau$  limit

$$\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \tilde{M}(A, \delta, \tau) = 0 \quad (1.249)$$

eliminates this term from Eq. (1.245), and one may state that  $A$  belongs to the domain of the steady state FR, which takes the form:

$$A - \delta \leq \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \ln \frac{\mu_{\infty}(\overline{\Omega}_{0,\tau}^{(0)} \in A_{\delta}^{+})}{\mu_{\infty}(\overline{\Omega}_{0,\tau}^{(0)} \in A_{\delta}^{-})} \leq A + \delta, \quad \forall \delta > 0 \quad (1.250)$$

As Eq. (1.249) is necessary for the validity of the SSFR in the form (1.250), it represents a condition that is surely verified when (1.250) holds. Its meaning is weaker than the decay of the correlations of the exponentials inside the conditional average in Eq. (1.245). It suffices that such correlations do not grow too fast. Also, it should be noted that unlike other derivations, here we speak of correlations with respect to the initial (known) probability distribution  $\mu^{(0)}$ , rather than the unknown steady state distribution  $\mu_{\infty}$ . In any event, although apparently minimal, this behaviour of correlations is not guaranteed, and SSFR for many systems do not follow Eq. (1.250), but some other form. For instance, the dissipated power  $\tilde{\epsilon}_{\tau}$  in gravitational bars subjected to feedback cooling has been observed to obey the following steady state FR [11]:

$$\rho(\tilde{\epsilon}_{\tau}) = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \ln \frac{\text{PDF}(\tilde{\epsilon}_{\tau})}{\text{PDF}(-\tilde{\epsilon}_{\tau})} = \begin{cases} 4\gamma\tilde{\epsilon}_{\tau}, & \tilde{\epsilon}_{\tau} < \frac{1}{3}; \\ \gamma\tilde{\epsilon}_{\tau} \left( \frac{7}{4} + \frac{3}{2\tilde{\epsilon}_{\tau}} - \frac{1}{4\tilde{\epsilon}_{\tau}^2} \right), & \tilde{\epsilon}_{\tau} \geq \frac{1}{3}. \end{cases} \quad (1.251)$$

It should now be clear that the difference between transient and steady state relations is substantial. In the first place, the transient ones hold identically whatever observation times one adopts, while the steady state one require long times, for a steady state to be reached, and then for correlations to behave reasonably well. When that does not happen, the steady state FR may not hold at all, or it may hold, but under a different form. These considerations have led to a general theory of response, not limited to small perturbations, and even capable of identifying conditions for its validity as single system, rather than as ensemble relations.<sup>22</sup>

---

<sup>22</sup>While thermodynamics needs single system relations, standard response theory yields ensemble relations; these become of interest when ensemble averages repeat the single system behaviour, but also when dealing with collections of independent small objects (with some proviso).

### 1.9.2 *t*-Mixing and a General Theory of Response

The necessary condition for the validity of the SSFR suggests a new kind of ergodic notion, concerning transient, rather than steady state probability distribution, that has been called *t*-mixing, and that can be expressed as [43, 45, 46]:

$$\lim_{t \rightarrow \infty} \left[ \langle (\mathcal{O} \circ S^t) \mathcal{P} \rangle^{(0)} - \langle \mathcal{O} \circ S^t \rangle^{(0)} \langle \mathcal{P} \rangle^{(0)} \right] = 0 \quad (1.252)$$

where  $\mathcal{O}$  and  $\mathcal{P}$  are two phase variables, and  $S^t$  is the evolution operator in phase space, up to time  $t$ . This equation resembles that of mixing, but it crucially differs from that, because averaging is performed with respect to a non invariant probability distribution,  $\mu^{(0)}$  say. Therefore, when this condition holds, the passage of time leads to a loss of memory about the initial macroscopic state. Differently, the same expression with an invariant probability distribution corresponds to the loss of correlations among microscopic events within a given stationary state. Taking  $\mathcal{P} = \Omega^{(0)}$  and observing that  $\langle \Omega^{(0)} \rangle^{(0)} = 0$ , Eq. (1.252) becomes:

$$\lim_{t \rightarrow \infty} \langle (\mathcal{O} \circ S^t) \Omega^{(0)} \rangle^{(0)} = 0 \quad (1.253)$$

Then, the condition under which

$$\int_0^\infty ds \langle (\mathcal{O} \circ S^s) \Omega^{(0)} \rangle^{(0)} \in \mathbb{R} \quad (1.254)$$

i.e. when  $\langle (\mathcal{O} \circ S^t) \Omega^{(0)} \rangle^{(0)}$  decays faster than  $1/t$ , is called  $\Omega t$ -mixing. This is important because some algebra proves that

$$\frac{d}{dt} \langle \mathcal{O} \rangle^{(t)} = \langle (\mathcal{O} \circ S^t) \Omega^{(0)} \rangle^{(0)} \quad (1.255)$$

which implies:

$$\langle \mathcal{O} \rangle^{(t)} = \langle \mathcal{O} \rangle^{(0)} + \int_0^t ds \langle (\mathcal{O} \circ S^s) \Omega^{(0)} \rangle^{(0)} \quad (1.256)$$

Thus, under  $\Omega t$ -mixing, the following exact response relation

$$\langle \mathcal{O} \rangle^{(\infty)} = \langle \mathcal{O} \rangle^{(0)} + \int_0^\infty ds \langle (\mathcal{O} \circ S^s) \Omega^{(0)} \rangle^{(0)} \quad (1.257)$$

holds. As in standard linear response theory, this formula expresses relaxation to a stationary state in the sense of ensembles, but is exact, and not limited to the linear regime. Furthermore, it applies to dynamical systems in general, and can be extended to stochastic processes as well. Finally, conditions can be given so that it expresses single system relaxation rather than ensemble response, which is important in view of the difficulties that affect the notion of ensembles in time dependent situations, [45].

## 1.10 From Ergodic Theory to Big Data

One of the main ideas behind big data and machine learning is that knowledge of the past (data) suffices to predict the future. So given time series of some kind of signal, one may think that there is an unknown underlying dynamical system that generates the signal. For simplicity, take the dynamics to be discrete in time:

$$\mathbf{x}_1 = \mathcal{S}\mathbf{x}_0 \quad (1.258)$$

$$\mathbf{x}_2 = \mathcal{S}\mathbf{x}_1 = \mathcal{S}^2\mathbf{x}_0 \quad (1.259)$$

⋮

$$\mathbf{x}_k = \mathcal{S}\mathbf{x}_{k-1} = \mathcal{S}^k\mathbf{x}_0 \quad (1.260)$$

Then, let  $A \subset \mathcal{M}$  be a measurable set in phase space, and denote by  $\tau_A(\mathbf{x})$  the recurrence time in that set, i.e. the shortest time after which a point of  $A$  returns to  $A$ :

$$\tau_A(\mathbf{x}) = \inf\{k \geq 1 : \mathbf{x} \in A, \text{ and } \mathcal{S}^k\mathbf{x} \in A\}, \quad (1.261)$$

If  $\mu$  denotes the invariant measure, which is the stationary probability distribution on  $\mathcal{M}$ , a theorem by Kac proves that the average recurrence time for ergodic system is given by:

$$\langle \tau_A \rangle = \frac{1}{\mu(A)}. \quad (1.262)$$

For systems with  $N$  degrees of freedom, whose range is  $O(L)$  for each component of  $\mathbf{x}$ , and for  $A$  of linear size  $\epsilon$ , one then obtains:

$$\langle \tau_A \rangle \sim \left(\frac{L}{\epsilon}\right)^N, \quad (1.263)$$

or

$$\langle \tau_A \rangle \sim \left(\frac{L}{\epsilon}\right)^D \quad (1.264)$$



if the steady state lives on a set of fractal dimension  $D$ . These times grow *exponentially* with  $N$ . In fact, this expresses in mathematical terms Boltzmann’s answer to Zermelo paradox: for a system of a macroscopic number of degrees of freedom, recurrence takes far longer than any physically imaginable time.

To build a model from known data, one looks for past states similar to the present one, and assumes that the future will approximately repeat what happened after the selected past state. This is the ancient method of analogues, that is translated in a modern mathematical language thanks to the notion of recurrence. In modern times, Maxwell noted that:

It is a metaphysical doctrine that from the same antecedents follow the same consequents. No one can gainsay this. But it is not of much use in a world like this, in which the same antecedents never again concur, and nothing ever happens twice. Indeed, for aught we know, one of the antecedents might be the precise date and place of the event, in which case experience would go for nothing. The metaphysical axiom would be of use only to a being possessed of the knowledge of contingent events, scintilla simplicis intelligentiæ degree of knowledge to which mere omniscience of all facts, scientia visionis, is but ignorance. The physical axiom which has a somewhat similar aspect is “from like antecedents follow like consequents”. But here we have passed from sameness to likeness, from absolute accuracy to a more or less rough approximation. There are certain classes of phenomena, as I have said, in which a small error in the data only introduces a small error in the result. Such are, among others, the larger phenomena of the Solar System, and those in which the more elementary laws in Dynamics contribute the greater part of the result. The course of events in these cases is stable.

Although he immediately added:

There are other classes of phenomena which are more complicated, and in which cases of instability may occur, the number of such cases increasing, in an exceedingly rapid manner, as the number of variables increases

In practice, the method of analogues assumes that the phenomena of interest are stable in Maxwell’s sense, and proceeds as follows. Given a sequence of past events

$$\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M \tag{1.265}$$

they are called “*analogues*” if they resemble each other in pairs with a certain degree of accuracy  $\epsilon$ , i.e. if:

$$|\mathbf{x}_i - \mathbf{x}_j| \leq \epsilon \tag{1.266}$$

Then, the approximate prediction that can be made is expressed by:

$$\mathbf{x}_{M+1} = \mathbf{x}_{k+1}, \quad \text{if } \mathbf{x}_k \text{ is an analogue of } \mathbf{x}_M \tag{1.267}$$

One then models the phenomenon defining a function  $\mathbf{f}$  such that sequences of states are approximated by  $\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k)$  within the chosen tolerance. Now, to find analogues one needs long series of data, as shown by Kac theorem: longer for higher accuracy or for larger space, and exponentially long in the number of degrees of freedom. For instance, even a relatively low accuracy such as  $\epsilon = L/10$  makes the task problematic,

because of the large dimensionality of the problems for which this approach is usually invoked, i.e. those that are too complex to be tackled in a theory that starts from an understanding of the fundamental mechanisms. The problem is enhanced in the case of chaotic dynamics, that exponentially magnify in time the deviations from the analog initial conditions. In principle,  $D$  equal to 6 or 7 is already out of reach in those cases.

One could add further considerations to refute the idea that the scientific method is made obsolete by the deluge of data. For instance, the tale attributed to Bertrand Russell of a turkey before Thanksgiving day, shows how naive induction is doomed to failure. The story goes like this: a turkey is fed every day. Every single feeding will firm up the bird's belief that it is the general rule of life to be fed every day by friendly members of the human race "looking out for its best interests," as a politician would say. On the afternoon of the Wednesday before Thanksgiving, something unexpected will happen to the turkey. It will incur a revision of belief ...

Nevertheless, there are successes in using data that cannot be denied, and that can certainly be integrated within a scientific framework, rather than opposing it; there still is a lot to understand about what can be extracted from data, and proceeding by examples is probably the most promising way to go. In fact, once again, it may be beneficial to distinguish sufficient from necessary conditions, noting that negative results, such as those about recurrence mentioned above do not prevent the method to succeed. For instance, in big data practice there might be implicit use of *a priori* knowledge, that needs to be made explicit.

## 1.11 Concluding Remarks

This paper gives an introduction to the field of nonequilibrium statistical mechanics, a branch of physics that is in rapid and even tumultuous growth, given the variety of possible applications, that go well beyond its original purpose: providing a microscopic foundation to equilibrium and nonequilibrium thermodynamics. Necessarily, we have just scratched the surface, starting from its foundations and its connection with thermodynamics. We have tried to explain why knowledge of atomic dynamics, nuclear properties, subnuclear ingredients are not enough to produce a macroscopic description of the world, and further ideas are needed when passing from one level of description to another. The problem is not that we should doubt that macroscopic objects are made of molecules, which are made of atoms, which are made of electrons and nuclei etc. The problem is what we know about these microscopic objects, which is then expressed by our theories; indeed, such theories do not always match in a smooth fashion [3]. The fact is that physics is a realm of measurements, and like our senses, measurement instruments only reveal certain aspects of the object of investigation, and is not uncommon that on different scales the resulting descriptions may even look contradictory: "*orthogonal*" to each other. It is one of the purposes of statistical physics to harmonize the different points of view, explaining what makes them different. This should also help us understand why our present knowledge of

atoms is not suitable e.g. to determine whether someone has got the flu, not to mention why one falls in love: a different perspective is surely needed for that.

Statistical mechanics in general makes use of concepts such as determinism and randomness, probabilities and material properties, and it daringly mingles them. This is necessary, given the breath of phenomena it intends to encompass. Understanding the foundation of statistical mechanics and, in particular, of its nonequilibrium section, helps its tools to extend their applicability beyond the well explored traditional grounds.

Indeed, understanding what makes probability material, in some sense, and what makes this identification fail in the cases in which the elementary constituents are atoms and molecules, sheds light on the other uses of probability, even those in which material properties are only vaguely present. These are most in need of investigation, and are indeed investigated in probabilistic terms, because no other tools seem to be available or equally satisfactory. At the same time probability may be used, but with little profit, if it is taken as an ethereal and mysterious entity.

Of course, statistical mechanics preserves all its interest in the study of the physical world, which is currently greatly developing, thanks to growing technological abilities in handling mesoscopic and microscopic spatiotemporal scales, but also in understanding geophysical scales, such as those concerning climate, oceans etc. Last but not least, even astrophysical and cosmological research has recently been approached from the nonequilibrium statistical mechanics point of view [47]. It will not be surprising in the future to see more and more fields of research open their doors to the nonequilibrium statistical mechanical approach.

**Acknowledgements** The author is indebted to Xiamen University and Huaqiao University for unique and generous hospitality. Deep gratitude is in order to Prof. Hong Zhao and his group, particularly Prof. Dahai He, for numerous deep and hearty scientific discussions, often accompanied by great food. The author thanks Prof. Giovanni Ciccotti, for reading this manuscript, and for the consequent intense, heated, exhilarating and enlightening discussions on foundational aspects of thermodynamics and statistical physics, which helped me shape the present text.

## Appendix 1: Exercise and Interpretation of Stochasticity

Consider a test particle of position  $x$  and velocity  $v$ , and a field particle of position  $x_1$  and velocity  $v_1$ . Suppose they obey the following set of differential equations:

$$\begin{cases} \dot{x} = v \\ \dot{v} = -\gamma v + x_1 \\ \dot{x}_1 = v_1 \\ \dot{v}_1 = -kx_1 \end{cases} \quad \text{with initial conditions} \quad \begin{cases} x(0) = 0 \\ v(0) = 0 \\ x_1(0) = x_{10} \\ v_1(0) = -kx_{10} \end{cases} \quad (1.268)$$

where  $k = \hat{k} + 1 > 1$  represents the effect of a spring and of the test particle on the field particle, and we assume that the initial conditions for the test particle are known,

while they are not known for the field particle. In this model, the field particle does not depend on the test particle, and its equations of motion can be readily solved:

$$x_1(t) = x_{10} \cos \sqrt{k}t + \frac{v_{10}}{\sqrt{k}} \sin \sqrt{k}t \quad (1.269)$$

$$v_1(t) = -\sqrt{k}x_{10} \sin \sqrt{k}t + v_{10} \cos \sqrt{k}t \quad (1.270)$$

Substituting  $x_1(t)$  in the equations of motion of the test particle, and solving, one obtains:

$$v(t) = \frac{x_{10}}{k + \gamma^2} \left[ \gamma \cos \sqrt{k}t + \sqrt{k} \sin \sqrt{k}t - \gamma e^{-\gamma t} \right] = \frac{x_{10}}{k + \gamma^2} f(t) \quad (1.271)$$

where the last equality defines the function  $f$ .

In the case in which the initial condition  $x_{10}$  is not known, the dynamics of the test particle remains also unknown. One may however know the probability distribution of its values. For instance, suppose  $x_{10}$  takes values in  $\{-1, 0, 1\}$ , with equal probability,  $1/3$ . Then we may compute the average of  $v$ , obtaining:

$$\mathbb{E}[v(t)] = \frac{1}{3}(-1 + 0 + 1) \frac{1}{k + \gamma^2} \left[ \gamma \cos \sqrt{k}t + \sqrt{k} \sin \sqrt{k}t - \gamma e^{-\gamma t} \right] = 0 \quad (1.272)$$

For the velocity autocorrelation function, note that:

$$v(t)v(t') = \left[ \frac{x_{10}}{k + \gamma^2} \right]^2 f(t)f(t') \quad (1.273)$$

which leads to

$$\mathbb{E}[v(t)v(t')] = \frac{1}{3}(1 + 0 + 1) \left[ \frac{1}{k + \gamma^2} \right]^2 f(t)f(t') = \frac{2}{3} \left[ \frac{1}{k + \gamma^2} \right]^2 f(t)f(t') \quad (1.274)$$

In summary, we have averaged the solution of the equations of motion of the test particle over the initial conditions weighted with their probability distribution. This can be interpreted as an average over an ensemble of identical and independent test particles, under the assumption that the dynamics of both test particles and field particles is deterministic and given by the corresponding equations of motion. This is the picture of Eq. (1.14), with  $N = 1$ . Clearly, there is a difference between trajectories that start from different initial conditions, and also a difference between the single trajectories and their averages. The same happens in the Brownian motion, of which this exercise is but an exaggeration. Indeed, one may legitimately maintain that for small particles in a liquid at room temperature, pressure etc, the classical mechanics description is accurate, the only difficulty being related to knowledge of the interaction potentials, and the huge number of degrees of freedom to consider. But this is a difficulty only in case quantitative explicit calculations are to be performed.

One may still conclude that such a detailed description is not necessary, and not even useful; in fact, as discussed earlier, the average picture is much more sensible than the detailed picture. In this case, we may say that we integrate the equations of motion first, and we average after:

$$\int_0^t \dot{v}(t'; x_{10}, v_{10}) dt' \mapsto \mathbb{E} \left[ \int_0^t \dot{v}(t'; x_{10}, v_{10}) dt' \right] = \int d\mu(x_{10}, v_{10}) \int_0^t dt' \dot{v}(t'; x_{10}, v_{10}) \tag{1.275}$$

where  $\mu$  is the probability distribution of the initial conditions of the test particle.

There is a practically almost equivalent point of view. One admits at start that only statistical properties make sense, and does not bother considering deterministic equations. After all, neither the deterministic nor the stochastic nature of a given phenomenon can be empirically validated; it is only a matter of deciding which of the available models better describes that phenomenon. Therefore, one may adopt the stochastic description not because of (possibly partial) ignorance of the initial state, but as a fully fundamental approach. In this case, one derives equations for probabilities and expectation values, and evolves them. Then it is like inverting the order of the integrals in Eq. (1.275):

$$\mathbb{E}_t [v] = \int_0^t ds \frac{d}{ds} \int d\mu_s(v) v = \int_0^t ds \dot{\mathbb{E}}_s [v] \tag{1.276}$$

where at time  $s$  the random variable  $v$  is distributed as prescribed by  $\mu_s$ , and no claim is made on the existence of  $\dot{v}$ . Within the deterministic view dealing with a finite number of particles  $N$  whose accelerations exist, averaging with respect to  $\mu_s$  amounts to sum over the  $N$  values of  $v$  and divide by  $N$ , therefore the two approaches are interchangeable.

## Appendix 2: Further Readings

It is impossible to do justice to so many different fields in a single paper. Therefore, we stop here, but we report a selection of readings, to complement the bibliography related to the content of the paper. Even this list is largely incomplete, but it provides further food for thought. I apologize for all unforgivable omissions. The list is presented in alphabetical order.

1. Paolo Adamo, Roman Belousov, and Lamberto Rondoni, Fluctuation-Dissipation and Fluctuation Relations: From Equilibrium to Nonequilibrium and Back, in A. Vulpiani et al. (eds.), Large Deviations in Physics, Lecture Notes in Physics 885, Springer-Verlag Berlin Heidelberg 2014
2. Ping Ao, Potential in stochastic differential equations: novel construction, Journal of physics A: mathematical and general 37 (3), L25

3. Urna Basu, Christian Maes, Nonequilibrium Response and Frenesy, *Journal of Physics: Conference Series* 638, 012001 (2015)
4. G. Boffetta, G. Lacorata, S. Musacchio, A. Vulpiani, Relaxation of finite perturbations: Beyond the fluctuation-response relation, *Chaos* 13 (2003) 806.
5. S. Bonella, G. Ciccotti and L. Rondoni Time reversal symmetry in time-dependent correlation functions for systems in a constant magnetic field, *EPL*, 108 (2014) 60004
6. H. B. Callen, *Thermodynamics and an introduction to thermostatistics*, Wiley (1985)
7. Li Can, Deng Wei-Hua and Shen Xiao-Qin, Exact Solutions and Their Asymptotic Behaviors for the Averaged Generalized Fractional Elastic Models, *Communications in Theoretical Physics* 62, 443 (2014)
8. Fabio Cecconi, Massimo Cencini and Angelo Vulpiani, Transport properties of chaotic and non-chaotic many particle systems, <https://doi.org/10.1088/1742-5468/2007/12/P12001>
9. C. Cercignani, Ludwig Boltzmann The Man Who Trusted Atoms, Oxford University Press (2006)
10. S Cervený, F Mallamace, J Swenson, M Vogel, L Xu, Confined water as model of supercooled water, *Chemical reviews* 116 (13), 7608–7625
11. Shunda Chen, Yong Zhang, Jiao Wang, Hong Zhao, Key role of asymmetric interactions in low-dimensional heat transport, *J. Stat. Mech.* (2016) 033205
12. Shunda Chen, Yong Zhang, Jiao Wang, and Hong Zhao, Diffusion of heat, energy, momentum, and mass in one-dimensional systems, *Phys. Rev. E* 87, 032153 – Published 25 March 2013
13. M Colangeli, L Rondoni, A Verderosa (2014). *Chaos Solitons Fractals* **64** 2
14. Leticia F. Cugliandolo, Jorge Kurchan, Pierre Le Doussal, Luca Peliti, Glassy behaviour in disordered systems with non-relaxational dynamics, *Phys. Rev. Lett.* 78, 350–353 (1997)
15. Sara Dal Cengio and Lamberto Rondoni, Broken versus Non-Broken Time Reversal Symmetry: Irreversibility and Response, *Symmetry* 2016, 8, 73; <https://doi.org/10.3390/sym8080073>
16. Christoph Dellago and Harald A. Posch, Realizing Boltzmann’s dream: computer simulations in modern statistical mechanics, in *Boltzmann’s Legacy* Giovanni Gallavotti Wolfgang L. Reiter Jakob Yngvason Editors
17. G Dematteis, T Grafke, E Vanden-Eijnden, Rogue waves and large deviations in deep sea, *Proceedings of the National Academy of Sciences* 115 (5), 855–860 (2018)
18. Abhishek Dhar, Heat Transport in low-dimensional systems *Advances in Physics*, Vol. 57, No. 5, 457–537 (2008)
19. Bertrand Duplantier, Brownian Motion, “Diverse and Undulating”, *Einstein 1905–2005, Poincaré Seminar 2005*, Birkhauser Verlag, Basel
20. Denis J. Evans, Gary P. Morriss, *Statistical Mechanics of Nonequilibrium Liquids*, Cambridge Univ. Press (2008)
21. G. Gallavotti, *Statistical Mechanics*, Springer (2009)
22. G. Gallavotti, *Nonequilibrium and Irreversibility*, Springer (2014)

23. H Ge, H Qian, Physical origins of entropy production, free energy dissipation, and their mathematical representations, *Physical Review E* 81 (5), 051133
24. H Ge, H Qian, Non-equilibrium phase transition in mesoscopic biochemical systems: from stochastic to nonlinear dynamics and beyond, *Journal of the Royal Society Interface* 8 (54), 107–116
25. Gnoli A, Puglisi A, Sarracino A, Vulpiani A (2014) Nonequilibrium Brownian Motion beyond the Effective Temperature. *PLoS ONE* 9(4): e93720. <https://doi.org/10.1371/journal.pone.0093720>
26. B Hu, B Li, H Zhao, Heat conduction in one-dimensional chains, *Physical Review E* 57 (3), 2992
27. K. Huang, *Statistical mechanics*, Wiley (1976)
28. G. Jacucci, Linear and nonlinear response in computer simulation experiments, *Physica A* 118 (1983) 157
29. Owen G. Jepps, Carlo Bianca and Lamberto Rondoni, Onset of diffusive behavior in confined transport systems, *CHAOS* 18, 013127 2008
30. D. Jou, J. Casas-Vazquez, G. Lebon, *Extended Irreversible Thermodynamics*, Springer (2001)
31. G. Lebon, D. Jou, J. Casas-Vázquez, *Understanding Non-equilibrium Thermodynamics*, Springer, Berlin (2008)
32. R. Klages, W. Just, C. Jarzynski (Eds). *Nonequilibrium statistical physics of small systems*. Wiley (2013)
33. R. Kubo, Statistical-mechanical theory of irreversible processes I, *J. Phys. Soc. Japan* 12, 570 (1957)
34. Kubo, R. (1986). Brownian Motion and Nonequilibrium Statistical Mechanics. *Science*, 233(4761), 330–334.
35. Toda, M., Kubo, R., Saito, N., *Statistical Physics I: Equilibrium Statistical Mechanics*, Springer (1983)
36. Lanford, Oscar E., III On a derivation of the Boltzmann equation. *International Conference on Dynamical Systems in Mathematical Physics (Rennes, 1975)*, pp. 117–137. *Asterisque*, No. 40, Soc. Math. France, Paris, 1976.
37. Lanford, Oscar E., III A derivation of the Boltzmann equation from classical mechanics. *Probability (Proc. Sympos. Pure Math., Vol. XXXI, Univ. Illinois, Urbana, Ill., 1976)*, pp. 87–89. *Amer. Math. Soc.*, Providence, R. I., 1977.
38. B Li, J Wang, Anomalous heat conduction and anomalous diffusion in one-dimensional systems, *Physical review letters* 91 (4), 044301
39. B Li, L Wang, G Casati, Thermal diode: Rectification of heat flux, *Physical review letters* 93 (18), 184301
40. Valerio Lucarini, Francesco Ragone and Frank Lunkeit, Predicting Climate Change Using Response Theory: Global Averages and Spatial Patterns, *J Stat Phys* (2017) 166: 1036
41. L. Luo and L.-H. Tang, Sub-diffusive scaling with power-law trapping times, *Chin. Phys. B* 23, 070514 (2014)
42. G. Morris, D. Evans, E. Cohen, H. van Beijeren, Linear response in phase-space trajectories to shearing, *Phys. Rev. Lett.* 62 (1989) 1579

43. L. Onsager, Reciprocal relations in irreversible processes. I, *Phys. Rev.* 37 (1931) 405–426.
44. Luca Peliti, *Statistical Mechanics in a Nutshell*, Princeton University Press (2003)
45. Hong Qian, A decomposition of irreversible diffusion processes without detailed balance, *J. Math. Phys.* 54, 053302 (2013)
46. Hong Qian a, Ping Ao b, Yuhai Tu c, Jin Wang, A framework towards understanding mesoscopic phenomena: Emergent unpredictability, symmetry breaking and dynamics across scales, *Chemical Physics Letters* 665 (2016) 153–161
47. Hong Qian, Shirou Wang, And Yingfei Yi, Entropy Productions In Dissipative Systems, *Proceedings Of The American Mathematical Society* <https://doi.org/10.1090/proc/14618>
48. Saar Rahav, Christopher Jarzynski, Nonequilibrium fluctuation theorems from equilibrium fluctuations. *New Journal of Physics.* 15. 125029 (2013)
49. David Ruelle, A review of linear response theory for general differentiable dynamical systems, *Nonlinearity* 22 (2009)
50. Udo Seifert, Stochastic thermodynamics, fluctuation theorems, and molecular machines, *Rep. Prog. Phys.* 75, 126001 (2012)
51. Herbert Spohn, Nonlinear fluctuating hydrodynamics for anharmonic chains, *Journal of Statistical Physics* 155, 1191–1227 (2014)
52. H. Spohn, *Large Scale Dynamics of Interacting Particles*, Springer (1991)
53. L. Tian, H. Ma, W. Guo and L.-H. Tang, Phase transitions of the q-state Potts model on multiply-laced Sierpinski gaskets, *Euro. Phys. J. B* 86, 197 (2013)
54. Hugo Touchette, The large deviation approach to statistical mechanics, *Physics Reports* 478 (2009) 1–69
55. M. Tuckerman, Lecture Notes on Statistical Mechanics. <http://www.nyu.edu/classes/tuckerman/stat.mech/lectures.html>.
56. C Villani, in *Boltzmann’s Legacy H-theorem and beyond: Boltzmann’s entropy in today’s mathematics*, Giovanni Gallavotti Wolfgang L. Reiter Jakob Yngvason Editors
57. Angelo Vulpiani, Fabio Cecconi, Massimo Cencini, Andrea Puglisi, Davide Vergni (Eds). *Lecture Notes in Physics; Large Deviations in Physics.* vol. 885, Springer (2014)
58. Y Wang, S Teitel, C Dellago, Melting of icosahedral gold nanoclusters from molecular dynamics simulations, *The Journal of chemical physics* 122 (21), 214722
59. Y Wang, ZC Tu, Efficiency at maximum power output of linear irreversible Carnot-like heat engines, *Physical Review E* 85 (1), 011127
60. Y Yuan, Q Wang, Y Shao, H Lu, T Li, A Gruverman, J Huang, Electric-field-driven reversible conversion between Methylammonium lead triiodide perovskites and lead iodide at elevated temperatures, *Advanced Energy Materials* 6 (2), 1501803
61. Yi Zhong, Yong Zhang, Jiao Wang, and Hong Zhao, Normal heat conduction in one-dimensional momentum conserving lattices with asymmetric interactions, *Phys. Rev. E* 85, 060102(R) (2012)



62. H Zhou, Network landscape from a Brownian particle's perspective, *Physical Review E* **67** (4), 041908
63. H Zhou, R Lipowsky, Dynamic pattern evolution on scale-free networks, *Proceedings National Academy of Sciences USA* **102** (29), 10052–10057
64. R. Zwanzig, *Nonequilibrium statistical mechanics*, Oxford University Press (2001)

## References

1. Lieb, E.H., Yngvason, J.: The physics and mathematics of the second law of thermodynamics. *Phys. Rep.* **310**, 1–96 (1999)
2. Koyré, A.: *Newtonian Studies*. Phoenix Books. The University of Chicago Press, Chicago (1968)
3. Chibbaro, S., Rondoni, L., Vulpiani, A.: *Reductionism, Emergence and Levels of Reality*. Springer, Berlin (2014)
4. Cantoni, G.: Su alcune condizioni fisiche dell'affinità e sul moto browniano. *Il Nuovo Cimento* **27**, 156–167 (1867)
5. Buenzli, P.R., Soto, R.: Violation of the action-reaction principle and self-forces induced by nonequilibrium fluctuations. *Phys. Rev. E* **78**, 020102 (2008)
6. Buenzli, P.R.: Fluctuation-induced self-force and violation of action-reaction in a nonequilibrium steady state fluid. *J. Phys. Conf. Series* **161**, 012036 (2009)
7. Pinheiro, M.J.: On Newton's third law and its symmetry-breaking effects. *Phys. Scr. bf* **84**, 055004 (2011)
8. Ivlev, A.V., Bartnick, J., Heinen, M., Du, C.-R., Nosenko, V., Löwen, H.: Statistical mechanics where Newton's third law is broken. *Phys. Rev. X* **5**, 011035 (2015)
9. Perrin, J.: *Les Atomes*. Alcan, Paris (1913)
10. Risken, H.: *The Fokker-Planck Equation*. Springer, Berlin (1989)
11. Bonaldi, M., et al.: Nonequilibrium steady-state fluctuations in actively cooled resonators. *Phys. Rev. Lett.* **103**, 010601 (2009)
12. Solon, A.P., Fily, Y., Baskaran, A., Cates, M.E., Kafri, Y., Kardar, M., Tailleur, J.: Pressure is not a state function for generic active fluids. *Nat. Phys.* **11**, 673–678 (2015)
13. De Groot, S.R., Mazur, P.: *Non-equilibrium Thermodynamics*. Dover Books on Physics (2011)
14. Das, Shibananda, et al.: Confined active Brownian particles: theoretical description of propulsion-induced accumulation. *New J. Phys.* **20**, 015001 (2018)
15. Sekimoto, K: *Stochastic Energetics*. Springer, Berlin (2010)
16. Kubo, R., Toda, M, Hashitsume, N.: *Nonequilibrium Statistical Mechanics*, Springer, *Statistical Physics II* (1991)
17. Laplace, P.S.: *Essai philosophique sur les probabilités*. H. Remy (1982)
18. Fermi, E.: *Thermodynamics*. Dover Books on Physics (2012)
19. Mañé, R.: *Ergodic Theory and Differentiable Dynamics*. Springer, Berlin (1987)
20. Sinai, YaG: Dynamical systems with elastic reflections. *Russ. Math. Surv.* **25**, 137 (1970)
21. Khinchin, Ya.A.: *Mathematical Foundations of Statistical Mechanics* Copertina Flessibile. Dover (1949)
22. Cercignani, C.: *The Boltzmann Equation and Its applications*. Springer, Berlin (1988)
23. Grad, H.: On the kinetic theory of rarefied gases. *Comm. Pure Appl. Math.* **2**, 331 (1949)
24. Ashcroft, N.W., Mermin, D.N.: *Solid State Physics*. Saunders College Publishing (1976)
25. Chen, F.F.: *Introduction to Plasma Physics and Controlled Fusion*. Springer, Berlin (2016)
26. Rondoni, L., Zweifel, P.F.: Collided-flux-expansion method for the transport of muonic deuterium in finite media. *Phys. Rev. A* **44**, 1104 (1991)

27. Melchionna, S., et al.: Hydrokinetic approach to large-scale cardiovascular blood flow. *Comput. Phys. Commun.* **181**, 462–472 (2010)
28. Falcioni, M., Palatella, L., Pigolotti, S., Rondoni, L., Vulpiani, A.: Initial growth of Boltzmann entropy and chaos in a large assembly of weakly interacting systems. *Physica A* **2007**(385), 170–184 (2007)
29. Saint-Raymond, L.: *Hydrodynamic Limits of the Boltzmann Equation*. Springer, Berlin (2009)
30. Kreuzer, H.J.: *Nonequilibrium Thermodynamics and Its Statistical Foundations*. Clarendon (1981)
31. Marini Bettolo Marconi, U., Puglisi, A., Rondoni, L., Vulpiani, A.: Fluctuation-dissipation: response theory in statistical physics. *Phys. Rep.* **461**, 111–195 (2008)
32. Boffetta, G., Lacorata, G., Musacchio, S., Vulpiani, A.: Relaxation of finite perturbations: beyond the fluctuation-response relation. *Chaos* **13**, 806 (2003)
33. Ruelle, D.: General linear response formula in statistical mechanics, and the fluctuation-dissipation theorem far from equilibrium. *Phys. Lett. A* **245**, 220 (1998)
34. Colangeli, M., Rondoni, L., Vulpiani, A.: Fluctuation-dissipation relation for chaotic non-Hamiltonian systems. *J. Stat. Mech.* **L04002** (2012)
35. Evans, D.J., Rondoni, L.: Comments on the entropy of nonequilibrium steady states. *J. Stat. Phys.* **109**, 895 (2002)
36. Bonetto, F., Kupiainen, A., Lebowitz, J.L.: Absolute continuity of projected SRB measures of coupled Arnold cat map lattices. *Ergod. Theory Dyn. Syst.* **25**, 59 (2005)
37. Cessac, B., Sepulchre, J.-A.: Linear response, susceptibility and resonances in chaotic toy models. *Physica D* **225**, 13 (2007)
38. Casimir, H.B.G.: On Onsager’s principle of microscopic reversibility. *Rev. Mod. Phys.* **17**, 343 (1945)
39. Coretti, A., Bonella, S., Rondoni, L., Ciccotti, G.: Time reversal and symmetries of time correlation functions. *Mol. Phys.* (2018)
40. Jepps, O.G., Rondoni, L.: Thermodynamics and complexity of simple transport phenomena. *J. Phys. A: Math. Gen.* **39**, 1311 (2006)
41. Giberti, C., Rondoni, L., Tayyab, M., Vollmer, J.: Equivalence of position-position auto-correlations in the Slicer Map and the Lévy-Lorentz gas. *Nonlinearity* **32**, 2302 (2019)
42. Evans, D.J., Cohen, E.G.D., Morriss, G.P.: Probability of second law violations in shearing steady states. *Phys. Rev. Lett.* **71**, 2401–2404 (1993)
43. Searles, Debra J., Rondoni, Lamberto, Evans, Denis J.: The steady state fluctuation relation for the dissipation function. *J. Statistical Phys.* **128**, 1337–1363 (2007)
44. Carberry, D.M., Baker, M.A.B., Wang, G.M., Sevicm, E.M., Evans, D.J.: An optical trap experiment to demonstrate fluctuation theorems in viscoelastic media. *J. Optics A Pure Appl. Opt.* **9**, S204 (2007)
45. Evans, D.J., Williams, S.R., Searles, D.J., Rondoni, L.: On typicality in nonequilibrium steady states. *J. Stat. Phys.* **164**, 842 (2016)
46. Jepps, O.G., Rondoni, L.: A dynamical-systems interpretation of the dissipation function, T-mixing and their relation to thermodynamic relaxation. *J. Phys. A Math. Theor.* **49**, 154002 (2016)
47. Baiesi, M., Burigana, C., Conti, L., Falasco, G., Maes, C., Rondoni, L., Trombetti, T.: Possible nonequilibrium imprint in the cosmic background at low frequencies. *Phys. Rev. Res.* **2**, 013210 (2020)

# Chapter 2

## On the Foundational Principles of Statistical Mechanics



Wei-Mou Zheng 

**Abstract** Several problems concerning the fundamental principles of statistical mechanics, including the canonical ensemble theory, the supporting set of an ensemble distribution function, the correspondence to thermodynamics, irreversibility, and time evolution, are briefly discussed.

**Keywords** Equilibrium statistical mechanics · Statistical ensemble · Entropy · Irreversibility

Here we first quote a few words from Chap. 7 of the book *Twentieth Century Physics* by Domb [1]. “In 1902 Gibbs introduced the concept of an ensemble [2], a collection of physical bodies following a statistical distribution . . . The properties of the body are calculated by taking a statistical average over the ensemble, and fluctuations above this average can be calculated using standard statistical procedures. . . . if the number of constituents,  $N$ , of the body becomes large, the fluctuations will decrease, and the averages can then be identified with thermodynamic properties in equilibrium. . . . Gibbs showed that by identifying  $-kT \ln Z_N$  (where  $Z_N$  is the partition function) with the free energy of Helmholtz,  $F(T, V, N)$ , results were obtained which coincided with thermodynamics.”

Gibbs introduced the grand canonical ensemble and chemical potentials, and used the grand partition function (GPF) to deal with the thermodynamics of heterogeneous substances. “Use of the GPF greatly simplifies the statistical mechanics of heterogeneous assemblies. . . . It took several decades for the scope and power of Gibbs approach to be appreciated.”

Gibbs “constructed a general theory of statistical mechanics which came to be seen as a foundation for all work in this field during the twentieth century. It is therefore appropriate to recognize him as a great pioneer of modern physics.” “A lucid series of lectures by Schrödinger [3] in Dublin in 1944 may well have been responsible for putting Gibbs ideas . . . as the basis of statistical mechanics . . .”

---

W.-M. Zheng (✉)  
Institute of Theoretical Physics, Academia Sinica, 100190 Beijing, China  
e-mail: [zheng@itp.ac.cn](mailto:zheng@itp.ac.cn)

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021  
X.-Y. Liu (ed.), *Frontiers and Progress of Current Soft Matter Research*,  
Soft and Biological Matter, [https://doi.org/10.1007/978-981-15-9297-3\\_2](https://doi.org/10.1007/978-981-15-9297-3_2)

Gibbs' ensemble concept, and the partition function to which the concept of ensemble naturally leads, can be regarded as an axiomatic representation of statistical mechanics. From the very beginning, he avoided the problem of the origin of statistical distributions. In this article several problems concerning the fundamental principles of statistical mechanics will be briefly discussed.

## 2.1 Statistical Laws

Consider a system of volume  $V$  containing  $N$  particles. Assuming it follows Hamiltonian dynamics, its state is described by the positions and momenta of the particles:  $(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N; \mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N) \equiv (\mathbf{r}^N, \mathbf{p}^N)$ , which is also called a microscopic conformation, or simply conformation. A conformation corresponds to a point in the  $6N$ -dimensional phase space spanned by  $\mathbf{r}^N$  and  $\mathbf{p}^N$ . Suppose the Hamiltonian of the system is  $\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N) = K(\mathbf{p}^N) + U(\mathbf{r}^N)$ , then the equations of motion are

$$\dot{\mathbf{r}}_j = \frac{\partial \mathcal{H}}{\partial \mathbf{p}_j}, \quad \dot{\mathbf{p}}_j = -\frac{\partial \mathcal{H}}{\partial \mathbf{r}_j}. \quad (2.1)$$

The change of conformations in time depicts a trajectory in the phase space, or a molecular orbit.

It is easy to write down the equations of motion, but it is impossible to solve or integrate the equations for any given initial conditions due to the extremely large number of degrees of freedom. However, this large number of degrees of freedom leads to totally new laws. As an object of the thermodynamics, a macroscopic system is always in some environment. Intrinsic (due to the unpredictability of chaotic dynamics) and extrinsic (from environmental noises) complications 'mix' up the molecular orbits. The picture of molecular orbit is destroyed, and an exact solution to the dynamics is no more necessary. New laws, which are statistical ones, appear in macroscopic systems. For example, the number of particles within a large enough volume element inside a container is rather steady. The laws result in the phenomenon observed in thermodynamics: a large system behaves quite simply and regularly, and can be characterized with a few independent variables. The language of molecular orbit is then replaced by that of probability distribution. A new kind of states, the state of statistical mechanics, can be defined as a distribution over the phase space specifying the probability for a system to appear around any point of the phase space. It should be emphasized that such statistical properties are independent of any details of the underlying microscopic laws. Whether they are classical or quantum, the framework of the theory of statistical mechanics does not change. States of thermodynamics are the macroscopic states in thermal equilibrium. They form the space of thermodynamical states, in which a path corresponds to a process of thermodynamics. As for quantities in thermodynamics, some can be obtained by averaging with the help of distributions of statistical mechanics; others are not averages and should be derived directly and entirely from the distributions, which is

worth noticing. It is not difficult to distinguish conformational states, thermodynamical states and states of statistical mechanics, but to fully understand the concepts behind them in analysing problems is often overlooked.

## 2.2 Canonical Ensemble

A state of statistical mechanics of a system means a distribution over the phase space. In general, the time scale of the dynamical evolution of molecular orbits is much shorter than that of the evolution of distributions. A system in statistical mechanics is always in an environment and of a large number of degrees of freedom. This makes exactly solving the dynamics of molecular orbits impossible and also unnecessary. On the same basis, a simple thermodynamical description exists for macroscopic equilibrium states. As a dynamical system, a microscopic system with a large number of degrees of freedom does not have any independent integrals of motions except for the Hamiltonian. In the ideal case, only the Hamiltonian, and only under a statistical average, is a conservative observable of the system. As a result of the wildly different time scales of the orbits and distributions, time plays a less important role in the equilibrium statistical mechanics. A principle of statistical mechanics implied by statistical laws states that a so-called canonical distribution depending only on the Hamiltonian exists for a system in thermodynamical equilibrium.

In 1902 Gibbs introduced the concept of ensemble, which is a collection of copies of a physical system under certain statistical distribution. Gibbs' theory of ensembles is axiomatic. The ensemble is not a physical thing. Gibbs did not attempt to derive distributions; the canonical ensemble is introduced by postulation. He merely designed a logical apparatus that produces relations which can be interpreted as analogues to the thermodynamical ones. The canonical distribution may be explained with the principle of maximum entropy, which claims that the probability distribution best representing the current state of knowledge, in the context of precisely stated testable information, is the one with largest entropy [4].

Suppose that the external parameters determining the Hamiltonian of a system, such as the number of particles and the volume, are given. In addition, the only testable information is the fixed value of

$$U = \int d\mathbf{r}^N d\mathbf{p}^N E(\mathbf{r}^N, \mathbf{p}^N) P(\mathbf{r}^N, \mathbf{p}^N) \equiv \langle E(\mathbf{r}^N, \mathbf{p}^N) \rangle, \quad (2.2)$$

where  $E$  is the energy of the system. According to the principle of maximum entropy, the distribution of the system should be

$$P(\mathbf{r}^N, \mathbf{p}^N) = Z(\beta, N, V)^{-1} \exp[-\beta E(\mathbf{r}^N, \mathbf{p}^N)], \quad (2.3)$$

where  $Z$  is a normalization factor or the partition function, and is defined as

$$Z(\beta, N, V) = \int d\mathbf{r}^N d\mathbf{p}^N \exp[-\beta E(\mathbf{r}^N, \mathbf{p}^N)]. \quad (2.4)$$

This probability of distribution  $P(\mathbf{r}^N, \mathbf{p}^N)$  under the only constrain of a fixed  $U$  maximizes the entropy

$$S = - \int d\mathbf{r}^N d\mathbf{p}^N P(\mathbf{r}^N, \mathbf{p}^N) \log P(\mathbf{r}^N, \mathbf{p}^N). \quad (2.5)$$

(When states are discrete, for a uniform distribution the entropy equals the logarithm of the number of states, and can thus be understood as the logarithm of the effective number of states.) Here we assume that the units have been chosen to make the Boltzmann constant be  $k_B = 1$ . The distribution parameter  $\beta = 1/T$  is the reciprocal temperature, i.e., the reciprocal of temperature  $T$ . As long as no confuse is caused, one may also refer to  $\beta$  as the temperature. (In statistical mechanics  $k_B T$  always appears as a combination.) The distribution described by Eq. (2.3) is called the canonical distribution or the Maxwell-Boltzmann distribution. The average energy  $U$  corresponds to the internal energy in thermodynamics, and can be derived from the partition function  $Z$  as

$$U = - \frac{\partial}{\partial \beta} \log Z(\beta, N, V). \quad (2.6)$$

Statistical mechanics studies systems in an environment, which may be a measuring apparatus or any systems in a given state, or a heatbath. A heatbath, as an ideal model for the environment, has as large a number as possible of degrees of freedom, but its dynamics and specific composition are unimportant. The essence of the model is that it is forever in thermodynamical equilibrium yet provides an energy exchange with systems inside it, and while doing this its own state never changes, or any changes are ignorable. The temperature is its most essential characterization. In a word, it is an energy pool with a constant temperature.

Note that in the definition (2.4) of the partition function that is often seen in textbooks no range for integration is specified. The range should be the supporting set of the distribution function. The determination of this supporting set, which depends on the system and problem under study, is the first step for calculations in statistical mechanics. The role played by this set in statistical mechanics has not been seriously surveyed, and will be further discussed later.

The testable information in a canonical ensemble restricts the average value of energy to the internal energy, but the energy in a canonical distribution fluctuates. While the internal energy is given by the derivative of the partition function  $Z$  with respect to temperature  $\beta$ , the energy fluctuation can also be derived from  $Z$ . The deviation of energy is

$$\begin{aligned}
\langle (\delta E)^2 \rangle &\equiv \langle (E - \langle E \rangle)^2 \rangle = \langle E^2 \rangle - \langle E \rangle^2 \\
&= \frac{1}{Z} \left( \frac{\partial^2 Z}{\partial \beta^2} \right)_{N,V} - \left[ \frac{1}{Z} \left( \frac{\partial Z}{\partial \beta} \right)_{N,V} \right]^2 = \left( \frac{\partial^2 \log Z}{\partial \beta^2} \right)_{N,V} = - \left( \frac{\partial U}{\partial \beta} \right)_{N,V}.
\end{aligned}
\tag{2.7}$$

By noticing the definition of the heat capacity at constant volume  $C_V = (\partial U / \partial T)_{N,V}$ , we obtain  $\langle (\delta E)^2 \rangle = \beta^{-2} C_V$ . Since the energy fluctuation is proportional to the square of temperature, it can be taken as a measure of temperature to reflect the strength of molecular orbit mixing. This outcome relates energy fluctuation with the response of energy to a change in temperature, and also embodies the linear response theory and fluctuation-dissipation theorem.

### 2.3 Microcanonical Ensemble

Canonical ensembles, also called *NVT* ensembles, describe a system with a constant number of particles  $N$  in a constant volume  $V$  and at a constant temperature  $T$ . Other ensembles such as the Gibbs ensemble, where the pressure instead of the volume is constant, can be derived from the canonical ensemble, and the equivalence among different ensembles can be proven. In the proof of the equivalence it is required that the sizes of both system and heatbath should be large. Although the volume in the Gibbs ensemble fluctuates, according to the central limit theorem, when the system size is large enough this fluctuation becomes negligibly small.

The microcanonical ensemble, also called *NVE* ensemble, refers to a system whose energy is limited in an infinitely narrow region centered at  $E$ . The microcanonical ensemble looks most simple, especially when setting principles for statistical mechanics, but it does not correspond to any real systems. Besides the inconvenience in calculation, there is an ambiguity relating the definition of entropy and temperature. Three types of entropy may be defined in the microcanonical ensemble in terms of the phase volume function  $V(E)$ , which counts the total number of states with energy less than  $E$  [5].

Defining the phase volume function in quantum mechanics is different from defining it in classical mechanics. It is necessary to introduce a smoothing function or normalized kernel  $f\left(\frac{H-E}{\omega}\right)$  over a width  $\omega$  centered at  $E$ . In quantum mechanics, density matrix  $\hat{\rho}$  plays the role of distribution:

$$\hat{\rho}(E) = \frac{1}{W} \sum_i f\left(\frac{H_i - E}{\omega}\right) |\psi_i\rangle \langle \psi_i|, \quad W = \sum_i f\left(\frac{H_i - E}{\omega}\right),$$

where  $H_i$  and  $|\psi_i\rangle$  are the Hamiltonian eigenvalue and eigenvector, respectively. When the limit  $\omega \rightarrow 0$  of the microcanonical ensemble is taken, original function of  $\delta(H - E)$  causes some trouble. If the width of energy surface is narrower than

the distance between energy levels, the count of the number of levels may be zero. Level degeneracy rarely occurs in a complex system, so the number of energy states changes discretely in energy. The derivative of the count with respect to energy becomes either zero or infinite. Thus, a smoothing function  $f$  over a finite width of energy surface is needed to avoid singularity, and the  $NVE$  ensemble becomes the  $NVE-\omega$  ensemble, with the distribution

$$\rho = \frac{1}{h^n C} \frac{1}{W} f\left(\frac{H-E}{\omega}\right), \quad W = \int d\mathbf{q}^N d\mathbf{p}^N \frac{1}{h^n C} f\left(\frac{H-E}{\omega}\right), \quad V(E) = \int_{H < E} d\mathbf{q}^N d\mathbf{p}^N \frac{1}{h^n C}.$$

Here  $C$  is an overcounting correction factor such as that used to correct for identical particles, and  $W$  is the effective volume of the expanded energy surface, and is given by  $W = \omega(dV/dE)$ .

As for the correspondence between the ensemble theory and thermodynamics, Boltzmann investigated only the ideal gas. A detail and thorough survey was accomplished by Gibbs. Three different types of entropies defined for the microcanonical ensemble are the Boltzmann entropy  $S_B$ , volume entropy  $S_v$  and surface entropy  $S_s$

$$S_B = k \log W = k \log(\omega dv/dE), \quad S_v = k \log v, \quad S_s = k \log(dv/dE) = S_B - k \log \omega.$$

Their associated temperatures are defined as  $1/T_v = dS_v/dE$  and  $1/T_s = 1/T_B = dS_s/dE = dS_B/dE$ . By using the volume entropy  $S_v$  and its associated  $T_v$ , it is possible to show exactly that

$$dE = T_v dS_v - \langle P \rangle dV,$$

is a close analogy to the first law of thermodynamics. A similar equation can be found for the surface entropy and Boltzmann entropy and their associated  $T$ ; however, pressure becomes a complicated quantity unrelated to an average. The microcanonical  $T_v$  and  $T_s$  are not entirely satisfactory in their analogy to temperature; for example, they do not indicate the direction of heat flow. A serious difficulty appears in the microcanonical ensemble when dealing with composite systems. If we denote the energies of system 1, 2 and their composite by  $E_1, E_2$  and  $E_{12} = E_1 + E_2$ , respectively, then in general,  $dS_{v1}/dE_1 = dS_{v2}/dE_2$  does not imply  $dS_{v1}/dE_1 = dS_{v2}/dE_2 = dS_{v,12}/dE_{12}$ . Only in a sense of averaging over the microcanonical ensemble of the composite system do we have  $dS_{v,12}/dE_{12} = \langle dS_{v1}/dE_1 \rangle_{E_{12}} = \langle dS_{v2}/dE_2 \rangle_{E_{12}}$ . Furthermore, in some systems the density of states is not monotonic in energy, and so they can change sign multiple times as the energy is increased.

Is it possible for the microcanonical ensemble to be equivalent to other ensembles such as the canonical ensemble. By comparing the microcanonical ensemble with the canonical ensemble, it is clear that if and only if  $E$  of microcanonical ensemble satisfies  $E = U$ , i.e., the internal energy of the system, are the two ensembles equivalent when fluctuations are ignored. In the literature only the overall non-equivalence of the microcanonical ensemble to other ensembles is emphasized [6]. The derivation of the microcanonical ensemble from the canonical ensemble would easily reveal



the crucial reason for this non-equivalence. The energy of a thermodynamical system fluctuates around its internal energy. When simulating Hamiltonian dynamics in molecular dynamics, the energy is conservative, but only when the total energy of the system is kept near its internal energy is the simulation meaningful. However, the internal energy of a system is usually unknown in advance, and uncontrollable. In order to meet a preset temperature, energy is often adjusted by scaling the average kinetic energy or by means of a simulated thermostat.

## 2.4 Phase Transition and Breaking of Ergodicity

The Ising model at temperature  $T = 0$  exhibits a state with the lowest energy, where all spins are parallel. Due to the ferromagnetic interaction  $J > 0$ , the same orientation of nearby spins favours energy, but disfavours entropy. However, at a low temperature the contribution of entropy to free energy is suppressed, which makes it possible for the spin orientation to be consistent across a macroscopic distance. That is, a long range order or coherent spin orientation occurs, and  $M \equiv \langle \sum_i \sigma_i \rangle$  does not vanish even in the absence of a field. This phenomenon is known as spontaneous magnetization. The highest temperature allowing this to appear is the critical temperature of the Ising model.

The Ising model in the absence of a field is symmetric with respect to the spin orientation being up and down. As a consequence of this symmetry, the exact calculation of  $M$  always results in zero since for every conformation with a total spin  $m = \sum_i \sigma_i$  being positive there must be a symmetric conformation of negative  $m$ , and they cancel each other [7]. Then how does the spontaneous magnetization occur? One mechanism is to introduce an ‘auxiliary field’, which approaches zero after an initial nonsymmetrical distribution is formed. The broken symmetry remains after the auxiliary field is removed. However, a more natural mechanism is the breaking of ergodicity, which is discussed below.

By summing over the Boltzmann factors of the conformations whose total spin  $m = \sum_i \sigma_i$  is set at a specified value  $\mu$ , we may define a ‘state-sum’  $y(\mu)$  and a ‘sub-free-energy’  $g(\mu)$  as follows:

$$y(\mu) = \sum_{m=\mu} e^{\beta E(\{\sigma_i\})}, \quad g(\mu) \equiv -T \log y(\mu).$$

Obviously, the partition function equals  $Y(T, h, N) = \sum_{\mu} y(\mu) = \sum_{\mu} e^{-\beta g(\mu)}$ , and  $y(\mu)/Y$  is the probability to observe the conformations with total spin  $\mu$ . From the above analysis of the spontaneous magnetization, function  $g(\mu)$  which we regard as a one-dimensional effective potential should behave as follows: it is a single well at high temperature, but becomes a double-well when the temperature is decreased below the critical temperature. An external field can make the potential asymmetric, such that the depths of the two wells will be unequal. Once a barrier appears in the

potential, its height should be  $\sim N^{(d-1)/d}$  by an estimation from the surface area of magnetic domain. Provided the height is scaled as some positive power of  $N$ , it approaches infinity at the thermodynamic limit. Besides the thermodynamic limit, an equilibrium state involves the dynamic limit that time approaches infinity. An infinitely high potential barrier would result in nonequivalence of the ordering of the two limits:

$$\lim_{t \rightarrow \infty} \lim_{V \rightarrow \infty}, \quad \lim_{V \rightarrow \infty} \lim_{t \rightarrow \infty}.$$

In the latter ordering, as long as the temperature is not extremely low, the system has a chance to visit both sides of the barrier, but this is not the case when the former ordering is taken. As for a real system, both its volume and the observation time are finite. Which ordering should be taken depends on the specific undergoing process. In a word, the phase transition accompanies the breaking of ergodicity.

In numerical simulations for gas or liquid, a uniform initial conformation is usually taken. If the parameters of the system correspond to the region of the gas-liquid phase transition, in general no gas-liquid coexistence can be seen, in analog to the case of the Ising model. If conformations of coexistence phase are wanted, say in simulation of interface behavior, quite different initial conditions must be chosen. For example, we can let all particles be locate in one side of the container, and then relax the system. The most remarkable difference between gas and liquid phases is their densities, especially at a temperature far from the critical point. When a proper cutoff radius is chosen, a criterion to associate a particle with a gas or liquid phase may be the number of particles inside the cutoff sphere centered at the reference particle. As the demonstration of the breaking of ergodicity, the two opposite kind of initial conditions mentioned above behave quite differently in evolution. A particle attributed to the gaseous state will remain in the gas phase for a long period, and so will a particle attributed to the liquid state remain in the liquid phase. When the size of the simulated system is not very large, the phenomenon of the ergodicity breaking is not so conspicuous. On increasing the size of the system, the lifetime of a particle in its particular gas or liquid phase will be greatly extended. It is worth further investigating whether and how a sharp change would occur.

## 2.5 Thermodynamic Limit and Supporting Set of Distribution

To prove conclusively the correspondence between statistical mechanics and thermodynamics, the existence of the thermodynamic limit is an essential prerequisite, and this existence, demonstrated *a posteriori*, depends on the nature of the Hamiltonian of the system [8–11]. Taking the grand canonical ensemble as an example, we need to prove the existence of the limit  $\lim_{V \rightarrow \infty} V^{-1} \log \Xi$ . Lee and Yang conducted a proof of the existence of this limit for a rather general Hamiltonian. S.-G. Ma suggested possible extensions of the proof and also pointed out the limitation of

the proof that the two-body potential used by Lee and Yang does not meet the most important interactions in physics such as the electromagnetic interaction [12].

As mentioned above, in most standard textbooks of statistical mechanics in the definition of the partition function (2.4) no region for integration, or the supporting set of the distribution function, is specified. When ergodicity is broken, as discussed in the last section, the region in the phase space visited by a physical system changes, and the supporting set of the distribution function changes correspondingly. The determination of the supporting set depends on the physical problem under study. For example, when dealing with a crystal, one has to consider firstly the symmetry of the lattice, which then specifies the supporting set of the distribution function.

Gibbs introduced the grand canonical ensemble to deal with chemical reaction systems. As an example, we may consider the reaction  $H + H \rightarrow H_2$  of the combination of two hydrogen atoms into a hydrogen molecule to compare quantum mechanics with statistical mechanics. A treatment in quantum mechanics starts from the Hamiltonian of two hydrogen atoms. The Born-Oppenheimer approximation is used to separate out the degrees of freedom for nuclei, and the Schrödinger equation of electrons is solved for electronic levels  $U_i(\mathbf{R})$  as a function of the nuclear distance  $\mathbf{R}$ . Bounded states are then interpreted as a hydrogen molecule. However, in the treatment of statistical mechanics, one has to include both the hydrogen atoms and the hydrogen molecules in the grand canonical ensemble at the beginning. From a pure atomic Hamiltonian no hydrogen molecules would be derived from statistical mechanics. The phase space for a pure atomic system of hydrogen atoms is different from that for a mixture of hydrogen atoms and molecules with regard to the breaking of ergodicity. The first order phase transition may be regarded as the simplest chemical reaction  $A \rightarrow B$ , and thus should be treated using grand canonical ensemble.

Whether in a simulation of molecular dynamics or in a Monte-Carlo simulation a prerequisite is that no change is made in the supporting set of distribution function. When applying accelerated or enhanced means for sampling, we must be careful. This is often overlooked in the literature.

## 2.6 Analogue to Thermodynamics: Heat and Free Energy

Thermodynamic is a branch of physics studying the relation between heat and temperature and between energy and work. It defines macroscopic variables such as internal energy, entropy and pressure, and describes universal relations among them without regard to any specific property of specific matter. Such general rules are represented by the four laws of thermodynamics.

The early motivation for thermodynamics was a desire to increase the efficiency of heat engines. The most fundamental concepts in thermodynamics are the system and environment, and the most fundamental subjects are thermodynamic state and process. A thermodynamic system is a macroscopic physical object. When a system is in a thermodynamical equilibrium with an environment under certain conditions,

its state is fully described by a few physical and chemical variables or state variables characterizing the macroscopic properties. The equilibrium state of a system is classified according to a finite number of independent state variables, which span a space of thermodynamic states. A point in the space corresponds to a thermodynamical state. State variables may be either extensive or intensive. Extensive variables are proportional to the size of the system, and are additive, while the latter are independent of the size. At least one extensive variable is needed to fix the size of the system. In fact, entropy does not meet the literal description of the definition of an extensive variable.

There are three kinds of thermodynamical variables [13]. The first are external parameters attributed to external environment and are independent of the interior of the system. The second are averages of dynamical functions of microscopic conformations. The third are typical, belong particularly to thermodynamics, have no microscopic meaning, can be grasped only in a macroscopic sense, and are associated totally with the distribution function. A representative of such variables is entropy. The equation of state describes the dependence among several state variables which are not completely independent of each other. A thermodynamic process describes a series of thermodynamical states, and a path in the space of thermodynamical states corresponds an equilibrium process. (A non-equilibrium or irreversible process has no path representation.) In dealing with thermodynamic processes, independent and dependent variables should be carefully specified. For example, in an isobaric process the pressure is selected as an independent variable, and held fixed. Thermodynamics is a phenomenological theory. For example, the equation of state satisfies some thermodynamical constrains, but its form depends on the property of the matter that the system consists of, so it is not derivable from thermodynamics. To obtain thermodynamical relations and compute thermodynamical quantities from molecular knowledge is the task of statistical mechanics.

The first law of thermodynamics defines the internal energy and divides it into work and heat, but only the internal energy is a state variable while heat and work are not and depend on the process. The commonest work is the mechanical work  $W_{\text{mech}} = -pdV$ , where  $p$  is pressure. When a system performs work on some external environment its volume increases, and  $W_{\text{mech}}$  is negative. The general form of work is  $W = \sum_i f_i dX_i$ , where  $f_i$  is a generalized force and  $X_i$  a generalized displacement.

In statistical mechanics the internal energy  $U$  is the average of energy:

$$U(\beta, N, V) = \int d\mathbf{r}^N d\mathbf{p}^N E(\mathbf{r}^N, \mathbf{p}^N) P(\mathbf{r}^N, \mathbf{p}^N), \quad (2.8)$$

where the probability density is  $P = e^{-\beta E}/Z$  for the canonical ensemble, while the temperature  $\beta$  and number  $N$  of particles are parameters and volume  $V$  is an independent variable. Therefore, the change in internal energy equals

$$dU = \int d\mathbf{r}^N d\mathbf{p}^N [P(\mathbf{r}^N, \mathbf{p}^N) dE(\mathbf{r}^N, \mathbf{p}^N) + E(\mathbf{r}^N, \mathbf{p}^N) dP(\mathbf{r}^N, \mathbf{p}^N)].$$

The change in Hamiltonian with volume resulting from the confinement of the container wall is  $dE/dV = -p$ , where  $p$  is independent of microscopic conformations, hence  $dE = -pdV$ . Thus, the first term inside the square bracket is the term describing work. From the expression of probability density we have

$$\beta E(\mathbf{r}^N, \mathbf{p}^N) = -\log Z(\beta, N, V) - \log P(\mathbf{r}^N, \mathbf{p}^N). \quad (2.9)$$

Thus, the second term inside the square bracket is

$$-\beta^{-1} \int d\mathbf{r}^N d\mathbf{p}^N \log P(\mathbf{r}^N, \mathbf{p}^N) dP(\mathbf{r}^N, \mathbf{p}^N) = -\beta^{-1} \int d\mathbf{r}^N d\mathbf{p}^N d[P(\mathbf{r}^N, \mathbf{p}^N) \log P(\mathbf{r}^N, \mathbf{p}^N)] = TdS. \quad (2.10)$$

This is just the heat absorption. (Here the fact that  $\log Z$  is a constant and probability  $P$  is normalized has been taken into account. As a result of this the term concerning  $\log Z$  makes no contribution.) We then arrive at the first law of thermodynamics:

$$dU = -pdV + TdS. \quad (2.11)$$

It is straightforward to extend this to the general form of work. If the number of particles is allowed to change, by introducing  $\mu = (\partial U / \partial N)_{\beta, V}$  we have

$$dU = -pdV + TdS + \mu dN, \quad (2.12)$$

where  $\mu$  is the chemical potential measuring the increment in internal energy by adding a particle. In summary, the change in internal energy due to the change in Hamiltonian is work, while the change in internal energy due to the change in distribution is heat.

The second law of thermodynamics is a fundamental postulate applicable in any phenomena involving heat, and is used to explain irreversible processes in nature. There are several equivalent formulations of the law. Clausius states that heat cannot spontaneously flow from cold regions to hot regions. In the language of statistical mechanics, the law may be retold as a minimum principle: the free energy of a system in a heat bath never increases in any spontaneous process. Here the reduction indicates the direction of process, and implies approaching equilibrium, which is beyond the extent reachable by the equilibrium statistical mechanics. The minimum principle of free energy claims only that amongst all states the equilibrium state has minimal free energy, with no reference to the direction of any process.

To simplify notation, we use summation instead of integration. The relative entropy of any two distributions  $\{Q_i\}$  and  $\{P_i\}$  is defined as

$$D(Q, P) = \sum_i Q_i \log(Q_i/P_i).$$

By means of  $\log x \leq 1 - x$ , and setting  $x = P_i/Q_i$ , it is easy to prove that  $D(Q, P) \geq 0$ , and the equality is valid only when  $\{P_i\}$  and  $\{Q_i\}$  are identical. Now denote the canonical distribution by  $P_i = Z^{-1} \exp(-\beta E_i)$ , and assume  $\{Q_i\}$  to be any other

distribution of the system. According to the non-negativeness of relative entropy, we have

$$\sum_i Q_i \log(Q_i/P_i) = \log Z + \beta \sum_i E_i Q_i + \sum_i Q_i \log Q_i = \log Z + \beta \langle E \rangle_Q - S_Q \geq 0, \quad (2.13)$$

where  $\langle X \rangle_Q$  is the average of random variable  $X$  under distribution  $Q$ , and  $S_Q$  the entropy associated with  $Q$ . In the case of identical  $P$  and  $Q$ , the equality implies

$$\log Z = -\beta \langle E \rangle_P + S[P] = -\beta U + S.$$

Thus,

$$F(\beta, N, V) \equiv F_P \equiv -\beta^{-1} \log Z(\beta, N, V) = U - TS, \quad (2.14)$$

which is just the Helmholtz free energy of thermodynamics. Accordingly the Helmholtz free energy associated with any distribution  $Q$  may be defined as

$$F_Q \equiv \langle E \rangle_Q + T \sum_i Q_i \log Q_i = U_Q - TS_Q. \quad (2.15)$$

Using (2.13), we have

$$F_Q \geq F(\beta, N, V). \quad (2.16)$$

That is, the equilibrium state has the minimal free energy. Incidentally, the inequality here is consistent with the principle of maximum entropy according to which the canonical ensemble has maximal entropy. Note that the temperature  $T$  present in the definition of free energy  $F_Q$  is the same as that in the equilibrium distribution  $P$ , or is the temperature attributed to the heatbath. That is to say, the free energy of an arbitrary state in statistical mechanics is defined only for a system in a heatbath.

By further using (2.12), we have

$$dF = -SdT - pdV + \mu dN. \quad (2.17)$$

which is the change in free energy in an equilibrium or reversible process, and may be denoted specially by  $(dF)_{\text{rev}}$ . Consider an isothermal process at constant volume from non-equilibrium to equilibrium. We then have  $(dF)_{\text{irrev}} < (dF)_{\text{rev}} = 0$ . This may further be written as  $dF \leq -SdT - pdV + \mu dN$ . We have thus derived the definition of free energy for a non-equilibrium state or any general state in statistical mechanics. However, it is impossible to deduce from equilibrium distributions the approach to equilibrium, which is a problem concerning the direction of thermodynamical processes.

## 2.7 Arrow of Time: Irreversibility

The macroscopic irreversibility has been visualized as the ‘arrow of time’, which governs all macroscopic phenomena. Microscopic reversibility and macroscopic irreversibility seem incompatible to each other, which has bothered physicists for ages. In the history of the development of statistical mechanics, non-equilibrium studies represented by molecular kinetics came first, followed by equilibrium studies. Many details about microscopic processes such as collision are involved in molecular kinetics. In his famous recursion theorem Poincaré proved that a finite mechanical system will return to a state arbitrarily close to its initial state [14]. This was like a bug that disturbed Boltzmann throughout his whole life. The axiomatic representation of equilibrium statistical distribution by Gibbs earned a great accolade (his monograph was highly praised by Poincaré just after publication.) In Gibbs’ theory the bug is unseen, but not removed. In fact, the microscopic reversibility is associated with the evolution of molecular orbits while the macroscopic irreversibility is associated with the evolution of distributions; the two need not contradict each other at all.

In the summer of 1953, Fermi, Pasta and Ulam (FPU) conducted a numerical simulation of a one-dimensional lattice of nonlinearly coupled oscillators on the MANIAC, an early computer just available then [15]. It was expected that energy would equalized among different modes, but only the phenomenon of Poincaré recursion was seen. This is called TPU paradox. (The FPU problem should be called the FPUT problem to acknowledge the contribution of Tsingou who programmed the MANIAC simulation.) The experiment designed by Fermi was to see molecular orbits, so of course only the Poincaré recursion was seen. If one wants to see the evolution of distribution, a set of initial states should be taken, and mixing among orbits allowed.

For exploring the connection between molecular orbits and distributions one approach starts from statistical mechanics, another from nonlinear dynamics. As for how to deal with nonlinear dynamics in the language of distribution, Kolmogorov claimed that only the invariant distribution which survives after the strength of the noise exerting on a system approaches zero is meaningful in physics. As a simple example of reversible dynamics, the baker map is defined as

$$(x', y') = \begin{cases} (2x, \frac{1}{2}y), & \text{for } 0 \leq x < \frac{1}{2}, \\ (2x - 1, \frac{1}{2}(y + 1)), & \text{for } \frac{1}{2} \leq x < 1. \end{cases}$$

It is worth demonstrating Kolmogorov’s idea with the baker map.

## 2.8 Time Evolution of the Distribution Function

As mentioned above, a state of statistical mechanics means a distribution function in the microscopic phase space. Statistical mechanics ultimately has to deal with the

time evolution of distribution and macroscopic thermodynamical variables, whose time scale is much greater than the microscopic dynamical time scale of molecular orbits (in  $6N$ -dimensional phase space). No time is involved in Gibbs' theory of ensemble, which specifies the equilibrium distribution function, assigns it with the minimal free energy, and makes no indication about approaching equilibrium. The Liouville equation is essentially equivalent to the dynamical equation of molecular orbits, and is not a proper starting point for statistical mechanics. The principle about the contrast between the time scales of orbits and distributions in Gibbs' theory is valid also for general nonequilibrium distributions. Continuing in the spirit of Gibbs' thinking, we have to give up the attempt to derive the time evolution equation of distributions from the dynamical equation of molecular orbits like the Liouville equation. A new principle is required to describe the evolution of distributions, in the same manner as in quantum mechanics where we do not attempt to derive Schrödinger equation from the Hamiltonian equation.

Necessary conditions for a legitimate time evolution equation of distributions include satisfying the conservation of probability, with the equilibrium distribution being its solution. It is most natural to take the master equation as a candidate. Let us consider the master equation for the time evolution of distributions specified by the transition rates  $T(\mathbf{z} \rightarrow \mathbf{z}')$ :

$$P_{t+1}(\mathbf{z}) = \int d\mathbf{z}' P_t(\mathbf{z}') T(\mathbf{z}' \rightarrow \mathbf{z}), \quad (2.18)$$

where  $\mathbf{z}$  is an abbreviated notation for conformation, while the transition probability  $T(\mathbf{z} \rightarrow \mathbf{z}')$  is completely determined by the Hamiltonian of the system and satisfies the following detailed balance condition:

$$P_{\text{eq}}(\mathbf{z}') T(\mathbf{z}' \rightarrow \mathbf{z}) = P_{\text{eq}}(\mathbf{z}) T(\mathbf{z} \rightarrow \mathbf{z}'), \quad \frac{T(\mathbf{z} \rightarrow \mathbf{z}')}{T(\mathbf{z}' \rightarrow \mathbf{z})} = \frac{P_{\text{eq}}(\mathbf{z}')}{P_{\text{eq}}(\mathbf{z})} = \frac{e^{-\beta H(\mathbf{z}')}}{e^{-\beta H(\mathbf{z})}}. \quad (2.19)$$

The above distribution dynamics guarantees that a system will approach equilibrium  $P_{\text{eq}}(\mathbf{z})$ . In summary, we propose a principle of statistical mechanics as follows:

The evolution dynamics for the distribution of a system in an environment or heatbath is described by the master equation (2.18) with transition rates satisfying the detailed balance condition (2.19).

Note that the detailed balance condition here is associated with the original  $6N$ -dimensional conformations; the condition need not be valid for a reduced distribution in a lower dimensional space.

The transition probability matrix  $T$  satisfies  $\int d\mathbf{z} T_k(\mathbf{z}' \rightarrow \mathbf{z}) = 1$ , exhibiting a right eigenvector  $\mathbf{1}$  (whose components are all equal to a positive number). If  $T$  is irreducible, according to the Perron-Frobenius theorem then its spectrum radius equals 1. The left eigenvector dual to  $\mathbf{1}$  is a non-negative vector, corresponding to the equilibrium distribution. The continuous state extension of the finite Markov chain is the theory of compact transition operators. Furthermore, discrete time may be extended to continuous time.



Transition probability  $T(\mathbf{z} \rightarrow \mathbf{z}')$  as an operator usually is not Hermitian. It is often convenient to introduce the following Hermitian operator  $t$ :

$$t(\mathbf{z}, \mathbf{z}') = t(\mathbf{z}', \mathbf{z}) \equiv \sqrt{\frac{P_{\text{eq}}(\mathbf{z})}{P_{\text{eq}}(\mathbf{z}')}} T(\mathbf{z} \rightarrow \mathbf{z}').$$

If the left and right eigenvectors of  $T$  for eigenvalue  $\lambda$  are respectively  $\Phi_\lambda(\mathbf{z})$  and  $\Psi_\lambda(\mathbf{z})$  with  $\Phi_1(\mathbf{z}) \equiv P_{\text{eq}}(\mathbf{z})$ , then eigenvector of  $t$  is  $\phi_\lambda(\mathbf{z}) = \Phi_\lambda(\mathbf{z})/\phi_1(\mathbf{z})$ , where  $\phi_1(\mathbf{z}) = \sqrt{\Phi_1(\mathbf{z})}$  belongs to eigenvalue 1, and all other eigenvalues are less than 1. Written in the Dirac notation, we have  $\Psi_\lambda(\mathbf{z}) \rightarrow |\Psi_\lambda(\mathbf{z})\rangle \equiv \phi_1(\mathbf{z})|\phi_\lambda(\mathbf{z})\rangle$  and  $\Phi_\lambda(\mathbf{z}) \rightarrow \langle\Phi_\lambda(\mathbf{z})| \equiv [\phi_1(\mathbf{z})]^{-1}\langle\phi_\lambda(\mathbf{z})|$ . The transition operator becomes

$$T(\mathbf{z} \rightarrow \mathbf{z}') = \sum_\lambda \lambda |\Psi_\lambda(\mathbf{z})\rangle \langle\Phi_\lambda(\mathbf{z}')|, \quad t(\mathbf{z} \rightarrow \mathbf{z}') = \frac{\phi_1(\mathbf{z}')}{\phi_1(\mathbf{z})} T(\mathbf{z} \rightarrow \mathbf{z}') = \sum_\lambda \lambda |\phi_\lambda(\mathbf{z}')\rangle \langle\phi_\lambda(\mathbf{z})|.$$

where eigenvectors are normalized by convention. The orthonormal relation is expressed as  $\langle\phi_\mu(\mathbf{z})|\phi_\nu(\mathbf{z})\rangle = \langle\Phi_\mu(\mathbf{z})|\Psi_\nu(\mathbf{z})\rangle = \delta_{\mu\nu}$ . An arbitrary distribution  $P(\mathbf{z})$  corresponds to  $\langle P(\mathbf{z})| = [\phi_1(\mathbf{z})]^{-1}\langle p(\mathbf{z})|$ .

A stochastic process is often used to describe a physical phenomenon for reduced degrees of freedom, relating to coarse-graining. Thermodynamics is almost only valid for static cases while high frequency processes mostly require some microscopic description. Although the Langevin equation is applicable to non-Markovian random forces, it is not convenient when dealing with nonlinear cases. On the other hand, the Fokker-Planck (FP) equation is applicable to nonlinear and nonsteady cases, but works only for white noise. In terms of the equilibrium solution  $P_{\text{eq}}$ , the FP equation may be written as

$$\frac{\partial P(u, t)}{\partial t} = \frac{\partial}{\partial u} \left\{ D(u) \left[ - \left( \frac{d}{du} \log P_{\text{eq}}(u) \right) + \frac{\partial}{\partial u} \right] \right\} P(u, t). \quad (2.20)$$

For a Brownian particle in potential  $U(x)$  the equilibrium solution is  $P_{\text{eq}}(x) = Ce^{-\beta U}$ , so

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial x} \left[ D(x) \left( \beta \frac{dU}{dx} + \frac{\partial}{\partial x} \right) \right] P.$$

Such an equation for distribution evolution meets the requirement of approaching equilibrium. The drift term that depends on the equilibrium solution means that the driving force will come from some effective field such as the mean field. Consider a dilute solution of polar molecules in a nonpolar solvent. The polarization  $\mathbf{P} = \mathbf{P}_d + \mathbf{P}_a + \mathbf{P}_e$  comes from the dipole orientation, distance and charge distributions, respectively. The last two terms have a lag in infrared and high-frequency optical region, so they can be absorbed into dielectric constant  $\epsilon_\infty$ , and only the orientation term needs to be considered. We suppose the density  $n$  to be uniform, so the equilibrium orientation distribution is

$$f_{\text{eq}}(\theta, \phi) = \frac{ne^{\beta\mu E \cos \theta}}{\int e^{\beta\mu E \cos \theta} d\Omega} \approx \frac{n}{4\pi} (1 + \beta\mu E \cos \theta), \quad \mathbf{P}_d = \langle \mu \cos \theta \mathbf{E} / E \rangle \approx \frac{1}{3} n \beta \mu^2 \mathbf{E}.$$

The FP equation describing the Brownian motion of dipole orientation is [16]

$$\frac{\partial f}{\partial t} = D \left\{ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left[ \sin \theta \left( \frac{\partial}{\partial \theta} + \beta\mu E(t) \sin \theta \right) f \right] + \frac{1}{\sin^2 \theta} \frac{\partial^2 f}{\partial \phi^2} \right\}.$$

Another example is the nuclear magnetic resonance. Quantities like the response function can be derived from statistical mechanics, related to correlation function, and manifested in a fluctuation-dissipation theorem. Equilibrium statistical mechanics provides static responses function while nonequilibrium statistical mechanics stresses dynamic responses.

Some coarse graining is necessary for statistical mechanics in order to reduce the dynamics of innumerable degrees of freedom to a stochastic evolution. The fundamental problem is to provide a bridge to sound mathematical logic, but this is yet to be solved. The equilibrium statistical mechanics established on the ergodicity theory is of a solid validity. Kubo has pointed out that nonequilibrium is far more difficult [16]. In the first place, its range is too wide, and has to be restricted. One of the two categories of methods includes the kinetic theory such as that based on Boltzmann's equation, which is applicable only for the cases when mean-free paths are long enough and field frequencies are sufficiently low though not limited to linear cases. Another category is the near-equilibrium theory, which relates nonequilibrium properties to equilibrium fluctuations, and is independent of coarse-graining. Van Kampen once severely criticized the linear response theory. He claimed that dynamical trajectories in phase space are essentially unstable and are very sensitive to perturbation, which makes the perturbation theory not meaningful. However, the linear response theory imposes a perturbation treatment only on the distribution instead of the orbit. Instability of trajectories will result in stability of distributions.

In summary, the crucial concept of statistical mechanics is the distribution function over the phase space. The key issue is to distinguish between the characteristic times of distributions and microscopic orbits. It is impossible to derive the evolution dynamics of distribution from dynamics of molecular orbits, and a new principle is required.

## References

1. Domb, C.: Thermodynamics and statistical mechanics (in equilibrium). In: Brown L, Pippard B, Pais A (eds) Twentieth Century Physics. AIP press (1995)
2. Gibbs, J.W.: Elementary Principles in Statistical Mechanics. Scribner, New York (1902)
3. Schrödinger, E.: Statistical Thermodynamics. Cambridge University Press, Cambridge (1946)
4. Jaynes, E.T.: Information theory and statistical mechanics I & II. Phys. Rev. **106**, 620–630; **108**, 171–190 (1957)
5. [https://en.wikipedia.org/wiki/Microcanonical\\_ensemble](https://en.wikipedia.org/wiki/Microcanonical_ensemble)

6. Hilbert, S., Hänggi, P., Dunkel, J.: Phys. Rev. E **90**, 062116 (2014)
7. Chandler, D.: Introduction to Modern Statistical Mechanics. Oxford University Press, Oxford (1987)
8. Bogoliubov, N.N.: J. Phys. USSR **10**, 257 (1946)
9. Van Hove, L.: Physica **15**, 951 (1949)
10. Yang, C.N., Lee, T.D.: Phys. Rev. **87**, 404 (1952)
11. Fisher, M.E., Ruelle, D.: J. Math. Phys. **7**, 260 (1966)
12. Ma, S.K.: Statistical Mechanics. World Scientific (1985)
13. Balescu, R.: Equilibrium and Nonequilibrium Statistical Mechanics. Wiley, London (1975)
14. Poincaré, H.: Acta Mathematica **13**, 1–270 (1890)
15. Fermi, E., Pasta, J., Ulam, S.: Studies of Nonlinear Problems. Document LA-1940. Los Alamos National Laboratory (1955)
16. Kubo, R., Toda, M., Hashitsume, N.: Statistical Physics II. Springer, Berlin (1998)

# Chapter 3

## Generalized Onsager Principle and Its Applications



Qi Wang 

**Abstract** We review the established nonequilibrium thermodynamical principle based on the Onsager linear response theory for irreversible processes, known as the Onsager principle for dissipative systems, firstly. We present it in two different forms: (1) the constructive formulation of the Onsager principle encompassing the kinetic equation, the reciprocal relation on the mobility and the maximum entropy principle; (2) the variational formulation in which the Onsager-Machlup action potential is maximized, or equivalently the Rayleighian is minimized in the isothermal case. Then, we generalize the Onsager principle in its constructive form to allow both reversible and irreversible processes to be modeled as well as mobility coefficient dependent on thermodynamical variables, which is termed the generalized Onsager principle (GOP). We carry out derivations of a plethora of thermodynamical and generalized hydrodynamical theories using the Onsager principle and the generalized Onsager principle to demonstrate their usefulness in establishing new models for nonequilibrium thermodynamical systems.

**Keywords** General Onsager principle · Onsager Machlup action potential · Generalized hydrodynamics · Non-equilibrium thermodynamical system

### 3.1 Introduction

Theories for equilibrium thermodynamics have been well developed based on the three fundamental thermodynamical laws: the first, the second and the third law of thermodynamics [8, 16, 17]. For nonequilibrium systems, however, theories are not quite so soundly developed in that there does not exist any physical laws like the three fundamental thermodynamical laws in nonequilibrium mechanics [3, 11]. For thermodynamical systems near equilibrium, the second law of thermodynamics in the form of the Clausius-Duhem inequality and the Onsager linear response theory

---

Q. Wang (✉)

Department of Mathematics, University of South Carolina, Columbia, SC 29208, USA  
e-mail: [qwang@math.sc.edu](mailto:qwang@math.sc.edu)

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021  
X.-Y. Liu (ed.), *Frontiers and Progress of Current Soft Matter Research*,  
Soft and Biological Matter, [https://doi.org/10.1007/978-981-15-9297-3\\_3](https://doi.org/10.1007/978-981-15-9297-3_3)

101

serve as corner stones for the development of many nonequilibrium theories for such systems [12–15]. In this chapter, we review the Onsager linear response theory for nonequilibrium systems, known as the Onsager principle [5], and make a contact with the second law of thermodynamics. The Onsager principle implies the second law of thermodynamics in the form of the Clausius-Duhem inequality. It is in fact more general in that it defines the dissipation mechanism that characterizes the constitutive relation for any matter systems. In the maximum Onsager-Machlup action potential principle, it provides a way to compute the dissipative force and match it up with the other non-dissipative forces. We then discuss how to generalize the Onsager principle to not only allow quasi-linear dependence of the mobility operator on thermodynamical variables in a functional way, but also to allow certain reversible processes directly related to the elastic force to be included in the model [21]. This allows one to apply the generalized Onsager Principle (GOP) to many complex systems beyond what the original Onsager principle was intended for. The organization of the chapter is as follows. In Sect. 3.2, we review the Onsager principle for dissipative systems in two equivalent yet distinct forms. In Sect. 3.3, we discuss the generalized Onsager principle. In Sect. 3.4, we derive a host of thermodynamical models including generalized hydrodynamical theories to show their variational and dissipative structures.

## 3.2 Onsager Principle for Dissipative Systems

We discuss the Onsager linear response theory for a purely dissipative system in details in this section. We proceed with it in a closed system first and then discuss it in an open system. We give two distinct formulations of the Onsager linear response theory together with the Onsager reciprocal relation in the name of the Onsager principle. The first formulation is constructive and the second is variational. Various applications of the Onsager principle in derivation of nonequilibrium thermodynamical models will be discussed as examples at the end of the section.

### 3.2.1 Constructive Onsager Principle

We consider a closed thermodynamical system not far from equilibrium  $\mathbf{x} = 0$ , whose state is described by a set of coarse-grained, thermodynamical variables  $\mathbf{x}(t) = (x_1, \dots, x_n)^T$  or fluctuations measured relative to equilibrium value  $\mathbf{x} = \mathbf{0}$ . Onsager states that entropy of the system  $S$ , which is assumed a function of  $\mathbf{x}$ , reaches its maximum value  $S_e$  at the equilibrium [13–15]. We expand the entropy function in its Taylor series at the equilibrium to arrive at the following approximation up to the quadratic term

$$S = S_e + \Delta S(\mathbf{x}) + o(\|\mathbf{x}\|^2), \quad \Delta S = -\frac{1}{2}\mathbf{x}^T \cdot \mathbf{H} \cdot \mathbf{x}, \quad (3.1)$$

where  $\mathbf{H}$  is the negative Hessian matrix of the entropy function, which is a symmetric and positive definite matrix ( $\mathbf{H} > 0$ ) at the maximum when  $S$  应为斜体 has the continuous second order derivatives. In fact, we can introduce a change of variables near the equilibrium so that the entropy function is exactly a quadratic form of a new set of thermodynamic variables  $\mathbf{x}'$  in a neighborhood of the maximum. So, without loss of generality, we assume the entropy function is a quadratic function of thermodynamical variable  $\mathbf{x}$  near the equilibrium. We note that this is plausible so long as we are interested in the near equilibrium behavior of the thermodynamical system. The probability density of thermodynamical variable  $\mathbf{x}$  near the equilibrium (when they are viewed as random fluctuations/variables) is related to  $\Delta S(\mathbf{x})$  via  $f(\mathbf{x}) = \frac{1}{Z} \exp[\Delta S(\mathbf{x})/k_B T]$ , where  $k_B$  is the Boltzmann constant,  $T$  the absolute temperature and  $Z$  is the partition function.

When the system deviates from equilibrium, spontaneous irreversible processes arise in response to the generalized thermodynamic force  $\mathbf{X}$  conjugate to  $\mathbf{x}$  defined by

$$\mathbf{X} = \left( \frac{\partial \Delta S}{\partial \mathbf{x}} \right) = -\mathbf{H} \cdot \mathbf{x}. \quad (3.2)$$

The force is created by the gradient in the entropy function. Its role is to drive the system in nonequilibrium back to its stable equilibrium state which we describe it as “relaxing” back to the equilibrium. The force vanishes at the equilibrium and this process is irreversible. The entropy production during the process, i.e., the entropy deviated from the equilibrium value, is given by

$$\Delta S = -\frac{1}{2} \mathbf{x}^T \cdot \mathbf{H} \cdot \mathbf{x} = \frac{1}{2} \mathbf{x}^T \cdot \mathbf{X}. \quad (3.3)$$

We note that the entropy production is the half of the inner product of the conjugate variable and the original thermodynamical variable (or fluctuation).

For a small deviation from the equilibrium, the system is assumed to be in the linear response regime, where the state  $\mathbf{x}(t)$  evolves according to the following kinetic equation

$$\dot{\mathbf{x}} = \mathbf{L} \cdot \mathbf{X}, \quad (3.4)$$

or, equivalently,

$$\mathbf{X} = \mathbf{L}^{-1} \cdot \dot{\mathbf{x}}, \quad (3.5)$$

where the coefficient  $\mathbf{L} = (L_{ij})$ , called the mobility, is an invertible matrix and its inverse is the friction coefficient matrix  $\mathbf{R} = \mathbf{L}^{-1}$ . Off-diagonal entries ( $L_{ij}$ ) and ( $R_{ij}$ ) are referred to as cross-coupling coefficients between different irreversible thermodynamical variables. Dynamics described by the kinetic equation is also known as the friction dynamics in Newtonian mechanics [4], which states that the velocity

of the motion in the dynamics is linearly proportional to the force acting on it. In the framework of the classical mechanics (Newtonian mechanics), this is equivalent to setting inertia to zero while retaining the friction force, which is assumed linear in  $\dot{\mathbf{x}}$ , and all the other external forces in the momentum balance equation. This system subject to friction is also known as the over-damped system in mechanics. Hence, the kinetic equation defines *relaxation dynamics* for how the nonequilibrium system relaxes back to the equilibrium in the linear response theory. Different mobility matrices define different dynamics although the entropy function of the system remains the same. It is therefore the crucial component in any nonequilibrium dynamical model.

Now that mobility defines actual dynamics of the system while relaxing back to equilibrium. We will relate the entropy production rate to the mobility coefficient explicitly in the irreversible process so that which dynamics the system adopts is fully determined by either specifying the mobility or the energy dissipation rate.

**Remark 3.2.1** The linear response adopted in this formulation is an assumption which dictates the development of the entire nonequilibrium theory. There is no reason why this cannot be modified to arrive at a truly nonlinear response and thereby yielding nonlinear, nonequilibrium response theories. However, this has not been explored in depth so far. For instance, the friction force might well be a nonlinear response function given by

$$\mathbf{X} = \mathbf{h}(\dot{\mathbf{x}}), \quad (3.6)$$

where  $\mathbf{h}(\dot{\mathbf{x}})$  is a nonlinear function of  $\dot{\mathbf{x}}$ . Assuming  $\mathbf{h}$  is invertible (for example,  $\mathbf{h}$  is monotonic), we have

$$\dot{\mathbf{x}} = \mathbf{h}^{-1}(\mathbf{X}), \quad (3.7)$$

where  $\mathbf{h}^{-1}$  is the inverse of  $\mathbf{h}$ . In order for the system to be dissipative, some constraint must be imposed on the function  $\mathbf{H}$ . To a large extent, how to proceed with this remains an open problem.

On the other hand, the mobility can be a function of the thermodynamical variable  $\mathbf{x}$ . So, the kinetic equation for the linear response can in fact be a nonlinear equation of  $\mathbf{x}$ . However, we hope one should keep in mind that the quasi-linear relation between the flux and the force in the linear response can deduce a nonlinear equation for the thermodynamical variables, but a nonlinear relation between  $\dot{\mathbf{x}}$  and  $\mathbf{x}$  may not always be a consequence of the linear response theory.

Under the condition that  $S$  is quadratic in  $\mathbf{x}$ , i.e., the sign of  $S$  remains invariant under a time-reversal operation, Onsager derived the well-known reciprocal relation

$$L_{ij} = L_{ji}, \quad (3.8)$$

and, consequently,  $R_{ij} = R_{ji}$ , from the microscopic reversibility, which states that for any  $t > 0$  and  $\tau$ ,

$$\langle \mathbf{x}(t)\mathbf{x}^T(t+\tau) \rangle = \langle \mathbf{x}(t+\tau)\mathbf{x}^T(t) \rangle, \quad (3.9)$$

where the ensemble is taken with respect to the probability density function  $f(\mathbf{x}(t))$  and  $f(\mathbf{x}(t+\tau))$ , respectively, and  $\mathbf{x}(t)\mathbf{x}^T(t+\tau)$  is the tensor product between the two vectors, known as the correlation matrix or tensor [12, 15].

Assuming  $\mathbf{x}(t)$  is governed by kinetic equation (3.4), we have for small  $|\tau|$

$$\mathbf{x}(t+\tau) = \mathbf{x}(t) + \mathbf{L} \cdot \mathbf{X}(t)\tau + O(\tau^2). \quad (3.10)$$

Substituting (3.10) into (3.9), we obtain

$$\langle (\mathbf{x}(t) + \mathbf{L} \cdot \mathbf{X}(t)\tau)\mathbf{x}^T(t) \rangle = \langle \mathbf{x}(t)(\mathbf{x}^T(t) + \mathbf{X}^T(t) \cdot \mathbf{L}^T\tau) \rangle + O(\tau^2). \quad (3.11)$$

We cancel the equal terms on both sides, divide the above equation by  $\tau$ , and then take limit  $\tau \rightarrow 0$  to obtain

$$\langle \mathbf{L} \cdot \mathbf{X}\mathbf{x}^T - \mathbf{x}\mathbf{X}^T \cdot \mathbf{L}^T \rangle = 0. \quad (3.12)$$

We next evaluate  $\langle \mathbf{X}\mathbf{x}^T \rangle$

$$\langle \mathbf{X}\mathbf{x}^T \rangle = \int \mathbf{X}\mathbf{x}^T f(\mathbf{x})d\mathbf{x}, \quad (3.13)$$

where  $f(\mathbf{x}) = \frac{1}{Z}e^{\Delta S(\mathbf{x})/k_B T}$ . Then,

$$\begin{aligned} \int \mathbf{X}\mathbf{x}^T f(\mathbf{x})d\mathbf{x} &= \frac{1}{Z} \int \frac{\partial \Delta S}{\partial \mathbf{x}} \mathbf{x}^T e^{\Delta S(\mathbf{x})/k_B T} d\mathbf{x} \\ &= \frac{1}{Z} \int \mathbf{x}^T \frac{\partial}{\partial \mathbf{x}} e^{\Delta S(\mathbf{x})/k_B T} d\mathbf{x} = -\frac{1}{Z} \int \mathbf{I} e^{\Delta S(\mathbf{x})/k_B T} d\mathbf{x} = -\mathbf{I}. \end{aligned} \quad (3.14)$$

It follows that

$$\mathbf{L} = \mathbf{L}^T, \quad (3.15)$$

i.e.,  $L$  is symmetric. It is worth emphasizing that this derivation does not require detailed knowledge of the irreversible process.

We next calculate the time rate of change of entropy,

$$\dot{S} = -\dot{\mathbf{x}}^T \cdot \mathbf{H} \cdot \mathbf{x} = \dot{\mathbf{x}}^T \cdot \mathbf{X} = \dot{\mathbf{x}}^T \cdot \mathbf{R} \cdot \dot{\mathbf{x}} = \mathbf{X}^T \cdot \mathbf{L} \cdot \mathbf{X}. \quad (3.16)$$

So, for a positive definite  $\mathbf{L} > \mathbf{0}$ , the entropy production rate is positive for any  $\mathbf{X} \neq 0$ . We identify  $\dot{\mathbf{x}}$  as the generalized flux given its time reversal property.

The maximum entropy assumption, the Onsager reciprocal relation and the kinetic equation defines entire near equilibrium dynamics for the thermodynamical system. These are collectively called the constructive Onsager principle. We next state the **Onsager principle** as follows:



1. the entropy is maximum at the equilibrium, i.e., the entropy production rate is nonnegative, which is ensured by the nonnegative definiteness of  $\mathbf{L}$ ;
2. dynamics near equilibrium is given by a linear response, i.e., kinetic equation (3.4);
3. the mobility coefficient matrix satisfies the reciprocal relation  $\mathbf{L} = \mathbf{L}^T$ .

We note that mobility in the original Onsager principle is assumed independent of  $\mathbf{x}$ . In modern applications, this has been generalized to allow  $L$  to be functional of  $\mathbf{x}$ . We will discuss this when we generalize the Onsager principle in the next chapter.

### 3.2.2 Onsager Principle Accounting for Inertia

The Onsager principle alluded to earlier applies to purely dissipative systems, in which inertia effect is not accounted for. We now turn to the system with non-negligible inertia and discuss how to incorporate the inertia effect into the Onsager principle. We consider a thermodynamical system in contact with its surrounding. Let  $U$  be its internal energy,  $E_k$  the kinetic energy corresponding to the inertia effect,  $Q$  its heat, and  $W$  the work the system does to the surrounding, the first law of thermodynamics states that

$$E_k + U = Q - W. \quad (3.17)$$

We note that the kinetic energy is a part of the total free energy which is independent of entropy. Let  $S$  be its entropy and  $T$  the absolute temperature defined from the second and the third law of thermodynamics, respectively. According to the second law, for a reversible process,

$$dS = \frac{dQ}{T}; \quad (3.18)$$

for an irreversible process on the other hand,

$$dS = \frac{dQ}{T} + \frac{dQ'}{T}, \quad (3.19)$$

where  $dQ'$  is the energy (in the form of heat) lost internally during the irreversible process. Recall that the Helmholtz free energy of the matter system is defined by

$$F = U - TS. \quad (3.20)$$

Hence,

$$d(E_k + F) = d(E_k + U) - TdS - SdT = -dW - dQ' - SdT. \quad (3.21)$$

For an isothermal system,  $dT = 0$ . Then,

$$d(E_k + F) = d(E_k + U) - TdS = -dW - dQ'. \quad (3.22)$$

The term on the right hand side is the energy loss to the surrounding and internally while the term on the left is the kinetic energy plus the Helmholtz free energy, termed the generalized free energy. The right hand side is termed energy loss or energy dissipation. It then follows that

$$\frac{d}{dt}(E_k + F) = -\frac{d(W + Q')}{dt}, \quad (3.23)$$

called the energy dissipation rate. This is consistent with the second law of thermodynamics.

By definition,

$$\frac{d}{dt}(E_k + F) = \frac{d}{dt}(E_k + U) - T \frac{dS}{dt} = \frac{d}{dt}(E_k + U) - T\mathbf{X}^T \cdot \mathbf{L} \cdot \mathbf{X}. \quad (3.24)$$

For a closed isothermal system, the total energy never changes, i.e.,  $\frac{d}{dt}(E_k + U) = 0$ . Then,

$$\frac{d}{dt}(E_k + F) = -T \frac{dS}{dt} = -T\mathbf{X}^T \cdot \mathbf{L} \cdot \mathbf{X}. \quad (3.25)$$

The energy dissipation rate is the negative of the entropy production rate. Let  $\mathbf{X}$  is the total generalized force. We adopt the kinetic equation from the liner response theory

$$\dot{\mathbf{x}} = \mathbf{L} \cdot \mathbf{X}. \quad (3.26)$$

From (3.61), we deduce

$$\frac{d}{dt}(E_k + F) = \left[ \frac{\partial}{\partial \mathbf{x}} F + \frac{d}{dt} \frac{\partial E_k}{\partial \dot{\mathbf{x}}} \right] \cdot \dot{\mathbf{x}} = -T\dot{\mathbf{x}}^T \cdot \mathbf{X}, \quad (3.27)$$

where  $E_k = \rho \|\dot{\mathbf{x}}\|^2/2$  is used. This implies

$$\mathbf{X} = -\frac{1}{T} \left[ \frac{\partial}{\partial \mathbf{x}} F + \frac{d}{dt} \frac{\partial E_k}{\partial \dot{\mathbf{x}}} \right] \quad (3.28)$$

So, in the thermodynamical system with inertia, the Onsager principle can be modified to accommodate the inertia. This version of the kinetic equation incorporates the force balance among the inertia, dissipative and the potential force, which generalizes the classical Onsager principle alluded to earlier.

In the constructive Onsager linear response theory, the entropy function is specified a priori and the dissipation functional is derived from the Onsager principle. In some cases however, we are given the system's dissipation function. In that case, how do we derive the kinetic equation for the system? We next give another equivalent approach to derive (3.4). I.e., we will show that the kinetic equation (3.4), can be derived via a variational principle provided the dissipation functional is known a priori. This is the variational version of the Onsager principle.

### 3.2.3 Variational Onsager Principle

We assume that we know the entropy production rate functional  $\dot{S}$  already and define

$$\Phi_S(\dot{\mathbf{x}}, \mathbf{x}) = \frac{1}{2} \dot{S} \quad (3.29)$$

as half of the entropy production rate functional depending on the generalized flux  $\dot{\mathbf{x}}$  and  $\mathbf{x}$ . Note that this is a quadratic function of  $\dot{\mathbf{x}}$  with a coefficient that can be functionals of  $\mathbf{x}$ . We then consider its Legendre transform with respect to  $\dot{\mathbf{x}}$  while treating  $\mathbf{x}$  as parameters:

$$O = \mathbf{X} \cdot \dot{\mathbf{x}} - \Phi_S = \dot{S} - \Phi_S. \quad (3.30)$$

Here  $\mathbf{X}$  is introduced as a conjugate variable to generalized flux  $\dot{\mathbf{x}}$ . The time integral of the functional is known as the Onsager-Matchlup action potential [18]:

$$OM = \int_{t_0}^t O[\dot{\mathbf{x}}(\tau), \mathbf{x}(\tau)] d\tau, \quad (3.31)$$

where  $t_0$  is a fixed initial time.

We note that  $\Phi_S$  is a convex function since  $\Phi_S(\dot{\mathbf{x}}, \mathbf{x})$  is as a negative definite quadratic functional of  $\dot{\mathbf{x}}$ . Viewing density function  $O$  as a function of  $(\mathbf{X}, \dot{\mathbf{x}})$ , we calculate its critical point with respect to  $\dot{\mathbf{x}}$  via variational principles to arrive at

$$\mathbf{X} - \frac{\partial}{\partial \dot{\mathbf{x}}} \Phi_S = \mathbf{X} - \mathbf{R} \cdot \dot{\mathbf{x}} = 0. \quad (3.32)$$

Equivalently,

$$\dot{\mathbf{x}} = \mathbf{L} \cdot \mathbf{X}. \quad (3.33)$$

This is kinetic equation (3.4). This indicates that the kinetic equation in Onsager linear response theory can be recovered from the maximum of the Onsager-Matchlup action potential. Of course, the Onsager-catchup action potential can be constructed conversely when the mobility is known.

In fact, if  $\mathbf{R} = \mathbf{L}^{-1} > 0$  and  $\mathbf{R}$  is symmetric,  $\mathbf{R} = (\mathbf{R}^{1/2})^2$ . Then,

$$\begin{aligned} OM &= \int_{t_0}^t [\mathbf{X} \cdot \dot{\mathbf{x}} - \frac{1}{2} \dot{\mathbf{x}}^T \cdot \mathbf{R} \cdot \dot{\mathbf{x}}] d\tau \\ &= \int_{t_0}^t [\mathbf{X} \cdot \dot{\mathbf{x}} - \frac{1}{2} \|\mathbf{R}^{\frac{1}{2}} \cdot \dot{\mathbf{x}}\|^2] d\tau \\ &= \int_{t_0}^t [\frac{1}{2} \|\mathbf{R}^{-1/2} \cdot \mathbf{X}\|^2 - \frac{1}{2} \|\mathbf{R}^{\frac{1}{2}} \cdot \dot{\mathbf{x}} - \mathbf{R}^{-1/2} \cdot \mathbf{X}\|^2] d\tau. \end{aligned} \quad (3.34)$$

The maximum of OM is achieved at (3.33) given the generalized force  $\mathbf{X}$ .

Notice that  $E_k$  is a quadratic function of  $\dot{\mathbf{x}}$  and  $F$  is a function of  $\mathbf{x}$ . So, when the inertia is considered,

$$\begin{aligned} \mathbf{X} \cdot \dot{\mathbf{x}} = \dot{S} &= -\frac{1}{T} \frac{d}{dt} (E_k + F) = -\frac{1}{T} \left[ \frac{\partial E_k}{\partial \dot{\mathbf{x}}} \ddot{\mathbf{x}} + \frac{\delta F}{\delta \mathbf{x}} \dot{\mathbf{x}} \right] \\ &= -\frac{1}{T} \left[ \frac{d}{dt} \frac{\partial E_k}{\partial \dot{\mathbf{x}}} + \frac{\delta F}{\delta \mathbf{x}} \right] \dot{\mathbf{x}}. \end{aligned} \quad (3.35)$$

This implies

$$\mathbf{X} = -\frac{1}{T} \left[ \frac{\delta F}{\delta \mathbf{x}} + \frac{d}{dt} \frac{\partial E_k}{\partial \dot{\mathbf{x}}} \right]. \quad (3.36)$$

Notice that we used the fact that for  $E_k = \frac{\rho}{2} \|\dot{\mathbf{x}}\|^2$ ,

$$\frac{d}{dt} \frac{\partial E_k}{\partial \dot{\mathbf{x}}} = \rho \frac{d\dot{\mathbf{x}}}{dt}, \quad (3.37)$$

which is the inertia force. Hence, the generalized force  $\mathbf{X}$  is proportional to the external forces minus the inertia force.

In practice, when inertia is considered, we use

$$O = \dot{S}(\mathbf{X}, \dot{\mathbf{x}}) - \Phi_S(\dot{\mathbf{x}}, \dot{\mathbf{x}}) = -\frac{d}{dt} \left[ \frac{1}{T} (E_k + F) \right] - \Phi_S \quad (3.38)$$

when differentiating it with respect to rate  $\{\dot{\mathbf{x}}\}$  while assuming  $\mathbf{X}$  is independent of  $\dot{\mathbf{x}}$ . It yields

$$\frac{\delta}{\delta \dot{\mathbf{x}}} (\dot{S} - \Phi_S) = \mathbf{X} - \mathbf{R} \cdot \dot{\mathbf{x}} = 0, \quad \Leftrightarrow \quad \dot{\mathbf{x}} = \mathbf{L} \cdot \mathbf{X}, \quad (3.39)$$

which is the kinetic equation in the Onsager principle.

Based on the above discussion, we state the variational form of the Onsager principle as follows:

### Onsager Principle in the variational form:

- Given a dissipation functional,  $2\Phi_S$ , as a quadratic function of  $\dot{\mathbf{x}}$ , we construct the Legendre transform of  $\Phi_S$  with respect to  $\dot{\mathbf{x}}$ :

$$O = \mathbf{X} \cdot \dot{\mathbf{x}} - \Phi_S = \dot{S} - \Phi_S, \dot{S} = -\frac{1}{T} \left[ \frac{\delta F}{\delta \mathbf{x}} + \frac{d}{dt} \frac{\partial E_k}{\partial \dot{\mathbf{x}}} \right] \dot{\mathbf{x}}. \quad (3.40)$$

The time integral of this functional is called the Onsager-Machlup action potential.

- Differentiating  $O$  with respect to  $\dot{\mathbf{x}}$ , one derives the kinetic equation for the system:

$$\frac{\delta O}{\delta \dot{\mathbf{x}}} = 0. \quad (3.41)$$

The kinetic equation is given explicitly by

$$\mathbf{X} = \frac{\delta \Phi_S}{\delta \dot{\mathbf{x}}} \Leftrightarrow \dot{\mathbf{x}} = L \left[ -\frac{1}{T} \left[ \frac{\delta F}{\delta \mathbf{x}} + \frac{d}{dt} \frac{\partial E_k}{\partial \dot{\mathbf{x}}} \right] \right]. \quad (3.42)$$

We note that the force  $\mathbf{X}$  is equal to the dissipative or frictional force. So, the Onsager principle gives one a way to calculate the dissipative force given the dissipation function and equates it to the other forces (conservative minus the inertia force) acted to the matter system. The Onsager principle in this variational form yields a force balance among all the forces acted to the system.

In a closed system, we have given two formulations of the Onsager principle: one in a constructive form, where the direct linear response theory is formulated, and the other in a variational form with a quadratic dissipation functional. These two formulations are equivalent. We next explore how we formulate the Onsager principle in an open system.

### 3.2.4 Onsager Principle in an Open System

For an open system under the isothermal condition, the rate of the total energy loss (internal+kinetic energy) is given by

$$\frac{d}{dt}(E_k + U) = -T \frac{dS^*}{dt}, \quad (3.43)$$

where  $S^*$  is the entropy lost to the surrounding. Here, we have to use the total entropy  $S_{total} = S + S^*$  in place of the entropy  $S$  in the closed system. We assume  $S_{total}$  is still a quadratic functional of  $\mathbf{x}$  and denote the conjugate force by  $\mathbf{X} = \frac{\partial(S+S^*)}{\partial \mathbf{x}}$  as well. We once again arrive at

$$\frac{d}{dt}(E_k + F) = \frac{d}{dt}(E_k + U - TS) = -T \frac{dS_{total}}{dt} = -T \mathbf{X}^T \cdot \mathbf{L} \cdot \mathbf{X}. \quad (3.44)$$

So, the energy dissipation rate is proportional negatively to the entropy production rate. It can be shown that the Onsager principle for irreversible thermodynamic processes is equivalent to the second law of thermodynamics in this case as well.

In an open system, we also have to determine the matter system more carefully. For example, when we consider a material domain  $\Omega$ , we must consider it as a closed domain with boundaries. The boundary of  $\Omega$ , denoted as  $\partial\Omega$ , itself constitutes a sub-matter system on which thermodynamics can take place. In this case, Onsager principle would have to be applied to both the interior of  $\Omega$  as well as its boundary  $\partial\Omega$ . This is especially important when we deal with free boundaries.

Consider the equivalent formulation in terms of the Onsager maximum action potential principle. For an open system, we must use the total entropy  $S_{total}$  in the Onsager-Machlup action potential density

$$O = \dot{S}_{total} - \Phi_S(\dot{\mathbf{x}}, \dot{\mathbf{x}}). \quad (3.45)$$

We assume  $S_{total}$  is a quadratic functional of  $\mathbf{x}$ . Then,  $\dot{S}_{total}$  is bilinear in  $\dot{\mathbf{x}}$  and  $\mathbf{X}$ . The Onsager variational principle states that for an open system, the state evolution equations can be obtained by differentiating Onsager-Machlup action potential density  $O$  with respect to flux  $\dot{\mathbf{x}}$  while viewing  $\mathbf{X}$  as independent of  $\dot{\mathbf{x}}$ . This principle serves as a general framework for describing irreversible processes in the linear response regime for an open system as well.

The Onsager variational principle is an extension of Rayleigh's principle of least energy dissipation and, naturally, it reduces to the latter in isothermal systems. In an isothermal system, the rate of entropy production given by the system to the environment can be expressed as

$$\dot{S}^* = -\frac{\dot{Q}}{T} = -\frac{\overline{\dot{U} + E_k}}{T}, \quad (3.46)$$

where  $T$  is the system temperature,  $\dot{Q}$  is the rate of heat transfer from the environment to the system, and  $\overline{\dot{U} + E_k}$  is the rate of change of the system total energy, with  $\dot{Q} = \overline{\dot{U} + E_k}$  according to the first law of thermodynamics, assuming no work is done during the process and the process is reversible. We note that The rate of energy change  $\overline{\dot{U} + E_k}$  must come from the surrounding in the open system since it is zero for a closed system. Using the Helmholtz free energy,  $F = U - TS$ , the Onsager-Machlup action potential density is rewritten into

$$O = \dot{S}_{total} - \Phi_S = \dot{S} + \dot{S}^* - \Phi_S = -\frac{\overline{\dot{F} + E_k}}{T} - \Phi_S, \quad (3.47)$$

where  $\dot{F} = \dot{U} - T\dot{S}$  in isothermal systems. We define the Rayleighian as

$$R = -TO = \dot{F}(\mathbf{x}) + \dot{E}_k(\dot{\mathbf{x}}, \dot{\mathbf{x}}) + \Phi_F(\dot{\mathbf{x}}, \dot{\mathbf{x}}), \quad (3.48)$$

where the dissipation function  $\Phi_F(\dot{\mathbf{x}}, \dot{\mathbf{x}}) = T\Phi_S(\dot{\mathbf{x}}, \dot{\mathbf{x}})$ . The maximization of the Onsager-Machlup action potential is equivalent to the minimization of the so-called time integral of the Rayleighian, which implies

$$\frac{\partial R}{\partial \dot{\mathbf{x}}} = 0 \Leftrightarrow -\frac{1}{T} \left[ \frac{\delta F}{\delta \mathbf{x}} + \frac{d}{dt} \frac{\partial E_k}{\partial \dot{\mathbf{x}}} \right] = L^{-1} \cdot x. \quad (3.49)$$

Here, we use

$$\overline{E_k + F} = \overline{E_k + U} - T\dot{S} = -T(\dot{S} + \dot{S}^*) = -T\dot{S}_{total}. \quad (3.50)$$

Notice that  $\dot{F}$  is linear in  $\mathbf{X}$  and  $\dot{\mathbf{x}}$ , while  $\dot{E}_k$  and  $\Phi_F$  are quadratic in  $\dot{\mathbf{x}}$ . If we substitute the kinetic equation ( $\mathbf{X} = \mathbf{L} \cdot \dot{\mathbf{x}}$ ) into  $\dot{F}$ , we obtain  $\overline{E_k + F} = -2\Phi_F$  under this dynamics or at the maximum of the Onsager-Machlup action potential density. This indicates that the time rate of change of the total free energy or the free energy dissipation rate is twice as large as the energy dissipation functional along the dynamical path of the nonequilibrium system. For isothermal systems, the Rayleighian is given explicitly by

$$R = \frac{\delta F}{\delta \mathbf{x}_i} \dot{\mathbf{x}}_i + \dot{E}_k(\dot{\mathbf{x}}, \dot{\mathbf{x}}) + \frac{1}{2} R_{T,ij} \dot{\mathbf{x}}_i \dot{\mathbf{x}}_j, \quad (3.51)$$

where  $\mathbf{R}_T = T\mathbf{R}$  is the rescaled friction coefficient, the generalized force is given by  $-\frac{\delta F}{\delta \mathbf{x}} - \frac{d}{dt} \frac{\partial E_k}{\partial \dot{\mathbf{x}}}$ . The first two terms in the right-hand side are  $\overline{E_k + F}$  and the third term is  $\Phi_F(\dot{\mathbf{x}}, \dot{\mathbf{x}})$ , which is in a quadratic form with the friction coefficients  $\mathbf{R}_T$  forming a symmetric and positive-definite matrix. Minimization of  $R$  with respect to rate  $\dot{\mathbf{x}}$  gives the kinetic equation

$$T\mathbf{X} = -\frac{\delta F}{\delta \mathbf{x}} - \frac{d}{dt} \frac{\partial E_k}{\partial \dot{\mathbf{x}}} = \mathbf{R}_T \cdot \dot{\mathbf{x}}, \quad (3.52)$$

which can be interpreted as a balance between the reversible force  $-\frac{\delta F}{\delta \mathbf{x}} - \frac{d}{dt} \frac{\partial E_k}{\partial \dot{\mathbf{x}}}$  and the dissipative or friction force linear in the flux.

If we represent the generalized force by  $T\mathbf{X}$ , the Rayleighian is given by

$$\begin{aligned} R &= -T\mathbf{X} \cdot \dot{\mathbf{x}} + \frac{1}{2} \dot{\mathbf{x}} \cdot \mathbf{R}_T \cdot \dot{\mathbf{x}} \\ &= -\frac{T}{2} \|\mathbf{R}^{-1/2} \cdot \mathbf{X}\|^2 + \frac{1}{2} \|\mathbf{R}^{1/2} \cdot \dot{\mathbf{x}} - (T\mathbf{R})^{-1/2} \cdot \mathbf{X}\|^2. \end{aligned} \quad (3.53)$$

The minimal Rayleighian corresponds to

$$T\mathbf{X} = \mathbf{L} \cdot \dot{\mathbf{x}}. \quad (3.54)$$

It is worth emphasizing that although the variational principle is equivalent to the kinetic equation combined with the reciprocal relation, the former possesses a notable advantage in some cases. The variational form allows flexibility in the choice of state variables. Once these variables are chosen, the conjugate forces are generated automatically via calculus of variations.

In summary, the extremal conditions on the Onsager-Machlup action potential density or the Rayleighian yields the force balance for the nonequilibrium system in isothermal case among the dissipative force, the inertia and the conservative forces. We next consider how to add external forces to the Onsager principle formulation when the force is nonconservative.

### 3.2.5 Effect of External Forces

If there exists an external body force to the system denoted as  $\mathbf{G}$ , we amend the Rayleighian by

$$R = \dot{\mathbf{x}} \cdot \left( \frac{\delta F}{\delta \mathbf{x}} - \mathbf{G} \right) + \dot{E}_k + \Phi_F. \quad (3.55)$$

Taking its derivative with respect to flux  $\dot{\mathbf{x}}$ , we obtain the force balance equation

$$T\mathbf{X} = \mathbf{R} \cdot \dot{\mathbf{x}} = -\frac{\delta F}{\delta \mathbf{x}} + \mathbf{G} - \frac{d}{dt} \frac{\partial E_k}{\partial \dot{\mathbf{x}}}. \quad (3.56)$$

Guided by this, we state the constructive Onsager principle for an open system with an external force  $\mathbf{G}$  is thus stated as follows:

1. the kinetic equation is given by  $\dot{\mathbf{x}} = \mathbf{L} \cdot \mathbf{X}$ , where generalized force is given by  $\mathbf{X} = \frac{1}{T} \left[ -\frac{\delta F}{\delta \mathbf{x}} + \mathbf{G} - \frac{d}{dt} \frac{\partial E_k}{\partial \dot{\mathbf{x}}} \right]$ ;
2. the reciprocal property  $\mathbf{L} = \mathbf{L}^T$  applies to the mobility;
3. the nonnegative definiteness of  $\mathbf{L}$  warrants positive entropy production.

Moreover, energy dissipation rate given by (3.61) is modified to

$$\frac{d}{dt} (E_k + F) = -T\mathbf{X}^T \cdot \mathbf{L} \cdot \mathbf{X} + \mathbf{G} \cdot \dot{\mathbf{x}} = \left[ -\frac{\delta F}{\delta \mathbf{x}} - \frac{d}{dt} \frac{\partial E_k}{\partial \dot{\mathbf{x}}} + \mathbf{G} \right] \cdot \dot{\mathbf{x}}. \quad (3.57)$$

The external force  $\mathbf{G}$  can be a conservative or a nonconservative force. In the former case, there exists a potential such that

$$\mathbf{G} = -\nabla h. \quad (3.58)$$

For example, the gravitational potential is  $h = g\mathbf{x} \cdot \mathbf{n}$ , where  $\mathbf{n}$  is the direction of the gravity and  $g$  is the gravitational acceleration. In this case, we effectively classify  $-\mathbf{G} \cdot \dot{\mathbf{x}}$  as a part of  $T\dot{S}^*$ .



### 3.2.6 Extension of Onsager Principle to Spatially Inhomogeneous Systems

In the presentation of the Onsager principle so far, we have assumed the entropy and the entropy production rate (analogously free energy and the energy dissipation rate) are functions of the thermodynamical variables. This applies to spatially homogeneous systems only. Realistically, most matter systems are spatially inhomogeneous. We can extend the Onsager principle to spatially inhomogeneous systems where potentials and mobility are functionals of the thermodynamical variables. Specifically, we assume the entropy, free energy and internal energy are functionals of the thermodynamical variables and the entropy production rate (or the energy dissipation rate) are functionals of the thermodynamical variables and quadratic functions of their corresponding invariant time derivatives. Under these assumptions, the thermodynamical principles mentioned above are still valid.

We generalize the dissipation functional to cases where the coefficients of the quadratic form in  $\dot{\mathbf{x}}$  depend on the thermodynamical variables and their spatial derivatives and/or integrals. Then, the kinetic equation derived from the Rayleighian is given by

$$T\mathbf{X} = \frac{\partial \Phi_F}{\partial \dot{\mathbf{x}}} = -\frac{\delta F}{\delta \mathbf{x}} + \mathbf{G} - \frac{d}{dt} \frac{\partial E_k}{\partial \dot{\mathbf{x}}}, \quad (3.59)$$

where the free energy functional, kinetic energy, and the dissipation functionals are given by

$$F[\mathbf{x}] = \int f(\mathbf{x}, \nabla \mathbf{x}, \dots) d\mathbf{r}, \quad E_k = \int \frac{\rho}{2} \|\dot{\mathbf{x}}\|^2 d\mathbf{r}, \quad \Phi_S = \frac{1}{2} \int \dot{\mathbf{x}} \cdot \mathbf{R} \cdot \dot{\mathbf{x}} d\mathbf{r}, \quad (3.60)$$

where  $f$  is the free energy density,  $\rho$  is the density of the matter system,  $r$  is the spatial variable,  $\mathbf{R}$  is the friction coefficient which can be functions of  $\mathbf{x}$  and its spatial derivatives. This yields all the transport equations for the system. Notice that this is a force balance equation that includes all the forces in the system: dissipative force  $-\frac{\delta \Phi_F}{\delta \dot{\mathbf{x}}}$ , elastic force  $-\frac{\delta F}{\delta \mathbf{x}}$ , external force  $\mathbf{G}$ , and the inertia force  $-\frac{d}{dt} \frac{\delta E_k}{\delta \dot{\mathbf{x}}}$ .

In practice, we either apply the Onsager principle directly to arrive at the constitutive relation by constructing the mobility matrix  $\mathbf{L}$  to define the kinetic equation:

$$\mathbf{X} = \mathbf{L} \cdot \dot{\mathbf{x}}, \quad (3.61)$$

or maximize the Onsager-Machlup action potential (or minimize the time integral of the Rayleighian) to arrive at the same relation when the dissipation functional  $\Phi_S$  is available. In the latter case, after we obtain the energy dissipation functional  $\Phi_S$  and the Helmholtz free energy  $F$ , the Rayleighian is defined by

$$R = \overline{E_k + F} - \mathbf{G} \cdot \dot{\mathbf{x}} + \Phi_F, \quad \phi_F = -T\Phi_S. \quad (3.62)$$

Minimizing the functional, we obtain

$$\mathbf{X} = \frac{1}{T} \left[ -\frac{\delta F}{\delta \mathbf{x}} + \mathbf{G} - \frac{d}{dt} \frac{\partial E_k}{\partial \dot{\mathbf{x}}} \right] = L^{-1} \cdot \dot{\mathbf{x}}, \quad (3.63)$$

where  $L$  is a functional of  $\mathbf{x}$ . In a nutshell, the Onsager maximum action potential principle basically calculates the dissipative force and balances it with all the other forces.

**Remark 3.2.2** The Onsager maximum action potential approach only works for a purely dissipative system. The constructive Onsager principle can be extended to account for reversible process as well, which we will consider in the generalized Onsager principle next.

In the following, we consider a series of well-known models and demonstrate how they can be derived using Onsager principles.

**Example 3.2.1** (Very viscous incompressible fluid model-Stokes equation) We consider a very viscous incompressible fluid in a domain  $\Omega$  where inertia can be effectively ignored. The dissipation functional is given by

$$\Phi = - \int_{\Omega} 2\eta \mathbf{D} : \mathbf{D} dx, \quad (3.64)$$

where  $\eta$  is the shear viscosity. Let  $\mathbf{x}(\mathbf{x}_0, t)$  is the position vector in the fluid at time  $t$  and  $\mathbf{x}_0$  is its Lagrange coordinate in a reference coordinate at  $t = 0$ . The fluid is incompressible means the Jacobian is a constant

$$J(\mathbf{x}, \mathbf{x}_0) = \left| \frac{\partial \mathbf{x}}{\partial \mathbf{x}_0} \right| = 1. \quad (3.65)$$

We differential this identity to obtain a constraint on the velocity

$$\nabla \cdot \mathbf{v} = 0. \quad (3.66)$$

The free energy for the system is consisted exclusively the constraint with a lagrange multiplier  $p$

$$F = - \int_{\mathbf{x}^{-1}(\Omega, t)} p J(\mathbf{x}, \mathbf{x}_0) d\mathbf{x}_0. \quad (3.67)$$

The time rate of change of the free energy is given by

$$\dot{F} = - \int_{\Omega} p \nabla \cdot \mathbf{v} d\mathbf{x} \quad (3.68)$$

The Rayleighian is then given by

$$R = \dot{F} + \frac{\Phi}{2}. \quad (3.69)$$

Applying the Onsager principle in variational form, we obtain

$$0 = -\nabla p + 2\eta \nabla \cdot \mathbf{D}. \quad (3.70)$$

This together with (3.66) constitute the Stokes equation for very viscous incompressible fluid flows.

**Example 3.2.2** (Incompressible viscous fluid model-Navier-Stokes equation) We consider an incompressible viscous fluid with a constant density  $\rho$  and incompressible constraint (3.66). The dissipation rate is given by (3.64). The Rayleighian is defined by

$$R = \int_{\Omega} \left[ \frac{d}{dt} \left( \frac{\rho}{2} \|\mathbf{v}\|^2 \right) - p \nabla \cdot \mathbf{v} \right] d\mathbf{x} + \Phi/2, \quad (3.71)$$

where  $\dot{\mathbf{v}} = \frac{d}{dt} = \left( \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \right) \mathbf{v}$  is the material derivative of  $\mathbf{v}$ . Applying the Onsager principle, we arrive at

$$0 = \rho \dot{\mathbf{v}} + \nabla p - 2\eta \nabla \cdot \mathbf{D}. \quad (3.72)$$

This together with (3.66) constitutes the Navier-Stokes equation for the incompressible viscous fluid flow.

**Example 3.2.3** (Compressible viscous fluid model) For a compressible viscous fluid, the free energy is given by

$$F = \int_{\Omega} f(\rho) d\mathbf{x}, \quad (3.73)$$

where  $f$  is the free energy density, a function of density  $\rho$ . The energy dissipation functional is given by

$$\Phi = - \int_{\Omega} [2\eta \mathbf{D} : \mathbf{D} + 2\nu (\nabla \cdot \mathbf{v})^2] d\mathbf{x}, \quad (3.74)$$

where  $\nu$  is the volumetric viscosity. The Rayleighian is defined by

$$\begin{aligned} R &= \int_{\Omega} \left( \frac{d}{dt} + \nabla \cdot \mathbf{v} \right) \left[ \frac{\rho}{2} \|\mathbf{v}\|^2 + f(\rho) \right] d\mathbf{x} + \Phi/2 \\ &= \int_{\Omega} \left[ \dot{\rho} \frac{\|\mathbf{v}\|^2}{2} + \rho \mathbf{v} \cdot \dot{\mathbf{v}} + \frac{\rho}{2} \|\mathbf{v}\|^2 \nabla \cdot \mathbf{v} + f'(\rho) \dot{\rho} + f(\rho) \nabla \mathbf{v} \right] d\mathbf{x} + \Phi/2, \end{aligned} \quad (3.75)$$

where  $\frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla$  is the material derivative.

We enforce mass conservation for the system

$$\dot{\rho} + \rho \nabla \cdot \mathbf{v} = 0. \quad (3.76)$$

The Rayleighian reduces to

$$R = \int_{\Omega} [\rho \mathbf{v} \cdot \dot{\mathbf{v}} + f'(\rho)(-\rho \nabla \cdot \mathbf{v}) + f(\rho) \nabla \mathbf{v}] d\mathbf{x} + \Phi/2. \quad (3.77)$$

Applying the Onsager principle, i.e, differentiating R with respect to  $\mathbf{v}$ , we have

$$-2\eta \nabla \cdot \mathbf{D} - 2\nu \nabla \cdot (\nabla \cdot \mathbf{v}\mathbf{I}) - \nabla(f + \frac{\rho}{2} \|\mathbf{v}\|^2) + \frac{d}{dt}(\rho \mathbf{v}) = 0. \quad (3.78)$$

The force balance equation reduces to

$$\frac{d}{dt}(\rho \mathbf{v}) = 2\eta \nabla \cdot \mathbf{D} + 2\nu \nabla \cdot (\nabla \cdot \mathbf{v}\mathbf{I}) + \nabla(f(\rho) - \rho f'(\rho)) \quad (3.79)$$

We define

$$p = \rho f'(\rho) - f(\rho) \quad (3.80)$$

as the Osmotic pressure. We arrive ta the compressible Navier-Stokes equation

$$\frac{d}{dt}(\rho \mathbf{v}) = 2\eta \nabla \cdot \mathbf{D} + 2\nu \nabla \cdot (\nabla \cdot \mathbf{v}\mathbf{I}) - \nabla p. \quad (3.81)$$

This together with (3.76) constitutes the model for the compressible viscous fluid flow.

**Example 3.2.4** (Incompressible binary viscous fluid model) We consider a binary viscous fluid of two viscous fluid components with the same densities. We denote the volume fraction of one fluid component as  $\phi$ . For the incompressible binary viscous fluid model in a domain  $V$ , the kinetic energy is given by  $E_k = \int_V \frac{\rho}{2} \|\mathbf{v}\|^2 d\mathbf{x}$ , the incompressibility condition is given by  $\nabla \cdot \mathbf{v} = 0$ , the dissipation functional is given by  $\Phi_F = \frac{1}{2} \int_V [\eta \|\nabla \mathbf{v}\|^2 + \gamma \|\mathbf{v}\|^2 + \dot{\phi} M^{-1} \dot{\phi} - p \nabla \cdot \mathbf{v}] d\mathbf{x}$ , where  $p$  is the pressure,  $M$  is the mobility operator, the free energy density is given by  $F[\phi] = \int_V f(\phi, \nabla \phi) d\mathbf{x}$ . The Rayleighian of the system is given by

$$R = \Phi_F + \overline{F + E_k}. \quad (3.82)$$

The kinetic equations for the incompressible fluid flows are given by

$$\begin{aligned}
\frac{\delta R}{\delta \mathbf{v}} &= -[\eta \nabla \cdot \nabla \mathbf{v} - \nabla p - \gamma \mathbf{v} - \rho \dot{\mathbf{v}}] + \nabla \phi M^{-1} \dot{\phi} = 0, \\
\frac{\delta R}{\delta p} &= \nabla \cdot \mathbf{v} = 0, \\
\frac{\delta R}{\delta \phi} &= \dot{\phi} + M \mu = 0,
\end{aligned} \tag{3.83}$$

where

$$\mu = \frac{\delta F}{\delta \phi}. \tag{3.84}$$

This system of equations can be simplified into

$$\begin{aligned}
\frac{d}{dt} \rho \mathbf{v} &= \eta \nabla \cdot \nabla \mathbf{v} - \nabla p - \gamma \mathbf{v} + \mu \nabla \phi, \\
\nabla \cdot \mathbf{v} &= 0, \\
\dot{\phi} + M \mu &= 0.
\end{aligned} \tag{3.85}$$

Here,  $(\dot{\bullet})$  must be the time invariant derivative, i.e., the material derivative.

**Remark 3.2.3** The mobility can be viewed as a differential operator. The time derivative is taken as the material derivative. If we set  $\phi = 1$  and the free energy as zero, we recover the transport equation for the incompressible viscous fluid. When the inertia term and the friction term are neglected, we recover the Stokes equation

$$\eta \nabla \cdot \nabla \mathbf{v} - \nabla p = 0. \tag{3.86}$$

When the inertia and the viscous stress is neglected, we recover the Darcy's law for fluid flows in the porous media

$$-\nabla p - \gamma \mathbf{v} = 0. \tag{3.87}$$

This equation can also be derived from the constructive Onsager principle directly by calculating the time rate of change of the total free energy and then apply the linear response theory. However, for the Stokes equation and the Darcy's law, the free energy simply consists of the Lagrange multiplier term for the constraint.

### 3.2.7 Lagrange Mechanics-A Complementary Formulation

We have learned that the Onsager principle provides a way for one to calculate the dissipative force and then put it against the other forces in a force balance equation for developing nonequilibrium models. Traditionally, there is yet another method

that one has used to derive mechanical models using Lagrangian mechanics for mechanical systems without dissipation. Let  $\mathcal{L}$  be the Lagrangian of the system defined by

$$\mathcal{L}[\mathbf{x}, \dot{\mathbf{x}}] = \int_{t_0}^{t_1} [E_k(\dot{\mathbf{x}}) - F[\mathbf{x}]] dt, \quad (3.88)$$

where  $[t_0, t_1]$  is a fixed time interval,  $E_k = \int \frac{\rho}{2} \|\mathbf{v}\|^2 d\mathbf{x}$  is the kinetic energy and  $F = \int f d\mathbf{x}$  is the free energy of the system. The variation of  $\mathcal{L}$  with respect to  $\mathbf{x}$  is given by

$$\frac{\delta \mathcal{L}}{\delta \mathbf{x}} = -\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}} + \frac{\partial \mathcal{L}}{\partial \mathbf{x}} = -\frac{d}{dt} \frac{\partial E_k}{\partial \dot{\mathbf{x}}} - \frac{\partial F}{\partial \mathbf{x}}. \quad (3.89)$$

In a system without an external nonconservative force, this is the total non-dissipative force. The above expression equals zero is known as the Euler-Lagrange equation for the system. This gives the governing system of equation for a nondissipative system. It does not apply to systems with dissipation.

If we balance this force with the other forces in a dissipative system including the dissipative force ( $\frac{\partial \Phi_F}{\partial \dot{\mathbf{x}}}$ ) and the nonconservative forces ( $-\mathbf{G}$ ), we end up with the force balance equation, i.e., the momentum balance equation

$$\frac{\delta \mathcal{L}}{\delta \mathbf{x}} + \mathbf{G} = T\mathbf{X} \quad (3.90)$$

for the dissipative system. This application extends the use of Lagrangian. The force  $T\mathbf{X}$  is the dissipative force, which is given by other means than the variation of the Lagrangian. If one has the ability to obtain the dissipative force by other means, the Lagrangian mechanics formulation can help us to obtain the non-dissipative forces. The dynamics of the system is determined through balancing the forces.

### 3.3 Generalized Onsager Principle

The Onsager reciprocal relation is valid for dissipative systems. In many nonequilibrium systems, the Onsager reciprocal relation does not hold or does not have to hold, for instance, conservative Hamiltonian systems, the active matter systems and viscoelastic fluid systems with microstructures etc. where irreversible and reversible processes coexist and intertwine with nonlocal interactions. For the systems, we have to extend the Onsager principle by allowing the mobility matrix to be non-symmetric and nonlinear and in the meantime the free energy to include nonlocal interactions. If we modify the mobility and free energy or entropy this way, we arrive at the generalized Onsager principle.

In the generalization of the Onsager principle, we assume the existence of a kinetic energy functional ( $E_k$ ), an entropy functional or a free energy functional ( $F$ ). Given a nonconservative external forces  $\mathbf{G}$ , the generalized force  $\mathbf{X}$  is given by

$$T\mathbf{X} = -\frac{\delta F}{\delta \mathbf{x}} + \mathbf{G} - \frac{d}{dt} \frac{\partial E_k}{\partial \dot{\mathbf{x}}}, \quad (3.91)$$

where  $T$  is the absolute temperature. The generalized Onsager principle states as follows.

1. There exists a mobility operator  $\mathbf{L}$  such that  $\dot{\mathbf{x}} = \mathbf{L}[\mathbf{x}] \cdot \mathbf{X}$ , where  $\mathbf{L}$  is a functional of  $\mathbf{x}$ ;
2. mobility operator  $\mathbf{L} = \mathbf{M}_{sym} + \mathbf{M}_{anti}$ , where  $\mathbf{M}_{sym} = \mathbf{M}_{sym}^T \geq 0$  and  $\mathbf{M}_{anti}^T = -\mathbf{M}_{anti}$ , may be functions of  $\mathbf{x}$ ;
3. the nonnegative definiteness of  $\mathbf{M}_{sym} \geq 0$  ensure the free energy is dissipative absent of any external forces in the isothermal cases

$$\frac{d}{dt}(F + E_k) = -T\mathbf{X}^T \cdot \mathbf{M}_{sym} \cdot \mathbf{X} + \mathbf{G} \cdot \dot{\mathbf{x}}, \quad (3.92)$$

or the entropy production is increasing

$$\frac{d}{dt} S_{total} = \mathbf{X}^T \cdot \mathbf{M}_{sym} \cdot \mathbf{X} - \frac{1}{T} \mathbf{G} \cdot \dot{\mathbf{x}}. \quad (3.93)$$

The antisymmetric part  $\mathbf{M}_{anti}$  corresponds to the reversible process that is non-energetic and not built in the Onsager-Machlup maximum action potential nor the Rayleighian so that the variational version of the Onsager principle can not be generalized. Thus, the generalized Onsager principle does not have an equivalent variational counterpart. By allowing antisymmetric mobility operators, we are able to handle energy conservative systems, like Hamiltonian systems. In the generalized Onsager principle, we basically say that the constitutive relation includes not only the dissipative force but also the nondissipative force!

**Remark 3.3.1** We have presented two distinct formulations of the Onsager principle for dissipative systems. One is in the form of the kinetic equation, maximum entropy principle and the Onsager reciprocal relation; and the other is in the form of the maximum Onsager-Machlup action potential or equivalently the minimum Rayleighian. In the presentation, we have clearly identified the thermodynamic variables or fluctuations away from the equilibrium as  $\mathbf{x}$  and assumed the entropy is given by a functional of these variables. Moreover, we assume the dissipation functional is given by a quadratic functional of flux variable  $\dot{\mathbf{x}}$ .

However, many dissipation functionals given for hydrodynamic theories in fluid systems are not given as functionals of  $\dot{\mathbf{x}}$ , but rather functionals of their spatial gradients in the Eulerian coordinate, like in the viscous fluid flows. For these systems, the corresponding generalized force is the stress tensor instead of the force. Normally,

there is no problem when identifying the thermodynamic variables and their conjugate variables if we apply the generalized Onsager principle directly to arrive at the kinetic equation. If one were to use the variational Onsager principle, one has to justify what is the thermodynamical variable whose variation needs to be considered. For the variational Onsager principle, regardless how the dissipation functional is defined, one should always minimize the Rayleighian with respect to  $\dot{\mathbf{x}}$ .

### 3.4 Applications of the Generalized Onsager Principle

The generalized Onsager principle provides a mathematical description for any thermodynamical systems near equilibrium. Here, we demonstrate how to apply it to derive thermodynamical models and hydrodynamical models for nonhomogeneous systems that satisfy the second law of thermodynamics.

#### 3.4.1 Dissipative Thermodynamical Models for Nonequilibrium Systems

We consider a nonequilibrium thermodynamical system with  $N$  internal variables to describe its state. The free energy of the system is given by

$$F = \int f(\{\phi_i\}, \{\nabla\phi_i\}, \{\nabla^2\phi_i\}, \dots) d\mathbf{x}, \quad (3.94)$$

where  $\phi_i, i = 1, \dots, N$  are the internal variables. The energy dissipation rate of the system is given by

$$\frac{dF}{dt} = \int (\mu_1, \dots, \mu_N) \cdot (\dot{\phi}_1, \dots, \dot{\phi}_N)^T d\mathbf{x}. \quad (3.95)$$

where  $\mu_i = \frac{\delta f}{\delta \phi_i}, i = 1, \dots, N$  are chemical potentials. We also consider the generalized inertia in the system corresponding to the kinetic energy

$$\int \frac{1}{2} \sum_{i,j=1}^N \phi_{i,t} \mathbf{C}_{ij} \phi_{j,t} d\mathbf{x}, \quad (3.96)$$

where  $\mathbf{C} > 0$  a constant ‘‘mass’’ matrix.

Generalized Onsager principle then implies

$$(\dot{\phi}_1, \dots, \dot{\phi}_N)^T = -\mathbf{M} \cdot [(\mu_1, \dots, \mu_N)^T + \mathbf{C} \cdot (\phi_{1,tt}, \dots, \phi_{N,tt})^T], \quad (3.97)$$



where  $\mathbf{M}$  is the mobility matrix. If we choose

$$\mathbf{M} = \sum_{k=0}^n \nabla^k \cdot \mathbf{L}_k \nabla^k, \quad (3.98)$$

where  $\mathbf{L}_k$  are symmetric tensor, the energy dissipation rate is given by

$$\begin{aligned} \frac{dF}{dt} = & - \int_{\Omega} \sum_{k=0}^n [\nabla^k [(\mu_1, \dots, \mu_m) + (\mathbf{C} \cdot (\phi_{1,t}, \dots, \phi_{N,t})^T)^T] \cdot \mathbf{L}_k \cdot \nabla^k [(\mu_1, \dots, \mu_m)^T + \\ & \mathbf{C} \cdot (\phi_{1,t}, \dots, \phi_{N,t})^T]] dx + \int_{\partial\Omega} g ds, \end{aligned} \quad (3.99)$$

where  $g$  is the energy dissipation rate across the surface. By choosing proper boundary conditions so that  $g = 0$  on the boundary, we arrive at the bulk energy dissipation rate equation/formula

$$\begin{aligned} \frac{dF}{dt} = & - \int_{\Omega} \sum_{k=0}^n [\nabla^k [(\mu_1, \dots, \mu_m) + (\mathbf{C} \cdot (\phi_{1,t}, \dots, \phi_{N,t})^T)^T] \cdot \mathbf{L}_k \cdot \nabla^k [(\mu_1, \dots, \mu_m)^T + \\ & \mathbf{C} \cdot (\phi_{1,t}, \dots, \phi_{N,t})^T]] dx. \end{aligned} \quad (3.100)$$

It is dissipative provided the quadratic form in the density is positive semi-definite. If  $\mathbf{C} = 0$ , this reduces to the dissipative model, also known as the gradient flow model. There are two well-known special cases:

- $n = 0$ , it yields the Allen-Cahn system.
- $n = 1$  and  $\mathbf{L}_0 = \mathbf{0}$ , it is the Cahn-Hilliard system.

This method can be used to derive thermodynamically consistent models for any material systems.

### 3.4.2 Gross-Pitaevskii Equations

Let  $u(\mathbf{x}, t)$  be a complex valued function. Consider the energy given by

$$F = \int [\nabla u \cdot \nabla \bar{u} - V_0(\mathbf{x})|u|^2 + V(|u|)] dx, \quad (3.101)$$

where  $V_0(\mathbf{x})$  is the trapping potential and  $V(|u|)$  is a nonlinear potential function for interactions. The energy dissipation rate is calculated as follows

$$\frac{dF}{dt} = \int [u_t (-\nabla^2 \bar{u} - V_0 \bar{u} + \frac{\partial V}{\partial |u|^2} \bar{u}) + \bar{u}_t (-\nabla^2 u - V_0 u + \frac{\partial V}{\partial |u|^2} u)] dx. \quad (3.102)$$

We set

$$u_t = M(-\nabla^2 u - V_0 u + \frac{\partial V}{\partial |u|^2} u). \quad (3.103)$$

If we impose that the energy is conserved, i.e.,  $\frac{dF}{dt} = 0$ . Then, it follows that

$$M = \pm i. \quad (3.104)$$

This gives rise to the Gross-Pitaevskii equation:

$$i u_t = \nabla^2 u + V_0 u - \frac{\partial V}{\partial |u|^2} u. \quad (3.105)$$

This is also known as the nonlinear Schrodinger equation. We can extend this derivation to a vector of complex valued function  $\mathbf{u}$  for multi-component Gross-Pitaevskii equations.

### 3.4.3 Generalized Hydrodynamic Theories

We consider a binary mixture of complex fluids consisting of polymers and solvent. The microstructure in the complex fluid is described by the low moments of a polymer distribution density function  $f(\mathbf{r}, \hat{\mathbf{v}}, t)$ :

$$c_n(\mathbf{r}, t) = \int_{\mathbf{S}} f(\mathbf{r}, \hat{\mathbf{v}}, t) d\hat{\mathbf{v}}, \quad \mathbf{p}(\mathbf{r}, t) = \frac{1}{c_n} \int_{\mathbf{S}} \hat{\mathbf{v}} f(\mathbf{r}, \hat{\mathbf{v}}, t) d\hat{\mathbf{v}}, \quad \mathbf{Q}(\mathbf{r}, t) = \frac{1}{c_n} \int_{\mathbf{S}} (\hat{\mathbf{v}}\hat{\mathbf{v}} - \frac{1}{d}) f(\mathbf{r}, \hat{\mathbf{v}}, t) d\hat{\mathbf{v}}, \quad (3.106)$$

where  $\mathbf{r}$  is the position vector,  $\hat{\mathbf{v}}$  is the orientation of the polymer,  $\mathbf{S}$  is the admissible space for  $\hat{\mathbf{v}} \in \mathbf{s}$  and  $d$  is the dimensionality.

We note that (i) hydrodynamics is described by low moments of a distribution function in any fluid systems; the momentum, density and energy are low moments of a distribution function; (ii) a generalized hydrodynamic model can be derived using either  $f$  or only the first a few low moments; (iii). hydrodynamics and microstructure couplings are via the low moments!

Let  $\rho$ ,  $\mathbf{v}$ ,  $\sigma$  and  $\mathbf{F}_e$  be the density, mass average velocity, total stress, and total body force of the mixture system. We then have the following conservation laws.

- **Mass conservation:**

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0. \quad (3.107)$$

- **Momentum conservation:**

$$\frac{\partial \rho \mathbf{v}}{\partial t} + \nabla \cdot (\rho \mathbf{v} \mathbf{v}) = \nabla \cdot \sigma + \mathbf{F}_e. \quad (3.108)$$

- Constitutive equation accounting for the microstructure of the system for  $\sigma$  and  $\mathbf{F}_e$ . These include:  $\sigma$  and  $\mathbf{F}_e$ , and transport equations for the internal variables  $(c, \mathbf{p}, \mathbf{Q})$ .

The singular limits of the momentum equation like the Stokes equation for very viscous mixtures are derived by taking the respective singular limits of the full equation.

Consider a mixture of complex fluids, whose micro-structures are described by the first 3 moments  $(c, \mathbf{p}, \mathbf{Q})$  of a microstructural distribution function and solvent with the total mass density  $\rho$  and mass averaged velocity  $\mathbf{v}$ . The mass conservation is given by (3.107). The number density of the microscopic constituent  $c$  in the complex fluid is described by the following equation

$$\partial_t c + \nabla \cdot (c\mathbf{v} + \mathbf{j}) = 0, \quad (3.109)$$

where  $\mathbf{j}$  is an extra diffusive flux.

The momentum equation is given by (3.108). The above equations need to be augmented by the constitutive equations. Namely, we need to derive evolution equations for  $c, \mathbf{p}, \mathbf{Q}$  and constitutive equation for flux  $\mathbf{j}$ , stress  $\sigma$  and  $\mathbf{F}_e$ .

### Constitutive equations:

We assume the free energy of the mixture system is given by

$$F = F[c, \nabla c, \mathbf{p}, \nabla \mathbf{p}, \mathbf{Q}, \nabla \mathbf{Q}] = \int_V f(c, \nabla c, \mathbf{p}, \nabla \mathbf{p}, \mathbf{Q}, \nabla \mathbf{Q}) dV, \quad (3.110)$$

where  $f$  is the free energy density per unit volume. The total free energy is defined by the sum of the kinetic energy and the free energy:

$$E^{total} = \int_V \left[ \frac{\rho}{2} \|\mathbf{v}\|^2 + f \right] dV. \quad (3.111)$$

The energy dissipation rate at a constant temperature  $T$  is calculated as follows

$$\begin{aligned} \frac{dE^{total}}{dt} &= \int_V \left\{ \frac{1}{2} \frac{\partial(\rho \|\mathbf{v}\|^2)}{\partial t} + \frac{\partial f}{\partial t} \right\} dV = \\ & \int_V [-\nabla \cdot (\frac{\rho \mathbf{v}}{2} \|\mathbf{v}\|^2) + \mathbf{v} \cdot (\nabla \cdot \sigma + \mathbf{F}_e) + \mu \frac{\partial c}{\partial t} - \mathbf{h} \cdot \frac{\partial \mathbf{p}}{\partial t} - \mathbf{G} : \frac{\partial \mathbf{Q}}{\partial t}] dV = \\ & \int_V [-\nabla \cdot (\frac{\rho \mathbf{v}}{2} \|\mathbf{v}\|^2) + \mathbf{v} \cdot (\nabla \cdot \sigma + \mathbf{F}_e) + \mu(\dot{c} - \nabla \cdot (c\mathbf{v})) - \mathbf{h} \cdot (\dot{\mathbf{p}} - \mathbf{v} \cdot \nabla \mathbf{p} - \Omega \cdot \mathbf{p}) - \\ & \mathbf{G} : (\dot{\mathbf{Q}} - \mathbf{v} \cdot \nabla \mathbf{Q} - \Omega \cdot \mathbf{q} + \mathbf{Q} \cdot \Omega)] dV = \\ & \int_{\partial V} [-\frac{\rho \mathbf{v}}{2} \|\mathbf{v}\|^2 + \mathbf{v} \cdot \sigma \cdot \mathbf{n}] ds + \int_V [(-\nabla \mathbf{v} : \sigma + \mathbf{h} \cdot \Omega \cdot \mathbf{p} + \mathbf{G} : (\Omega \cdot \mathbf{Q} - \mathbf{Q} \cdot \Omega)) + \\ & (\mathbf{v} \cdot \mathbf{F}_e - \mathbf{v} \cdot \nabla c + \mathbf{h} \cdot \mathbf{v} \cdot \nabla \mathbf{p} + \mathbf{G} : \mathbf{v} \cdot \nabla \mathbf{Q}) + \mu \dot{c} - \mathbf{h} \cdot \dot{\mathbf{p}} - \mathbf{G} : \dot{\mathbf{Q}} - r\hat{\mu}] dV = \\ & \int_{\partial V} [-\frac{\rho \mathbf{v}}{2} \|\mathbf{v}\|^2 + \mathbf{v} \cdot \sigma \cdot \mathbf{n} - \mathbf{n} \cdot \mathbf{j} \mu] ds + \int_V [-\mathbf{D} : \sigma^s + \nabla \mu \cdot \mathbf{j} - \mathbf{h} \cdot \dot{\mathbf{p}} - \mathbf{G} : \dot{\mathbf{Q}}] dV, \end{aligned} \quad (3.112)$$

where  $\mu = \frac{\delta F}{\delta c}$ ,  $\mathbf{h} = -\frac{\delta F}{\delta \mathbf{p}}$ , and  $\mathbf{G} = -\frac{\delta F}{\delta \mathbf{Q}}$  are the variation of the free energy with respect to the three internal variables,  $\sigma^a$  is the antisymmetric part of the stress,  $\sigma_e$  is the Ericksen stress,  $\sigma_s$  is the symmetric part of the stress,

$$\begin{aligned}
\sigma_{\alpha\beta}^s &= \sigma_{\alpha\beta} - \sigma_{\alpha\beta}^e - \sigma_{\alpha\beta}^a, \\
\sigma_{\alpha\beta}^a &= \frac{1}{2}(p_\alpha h_\beta - p_\beta h_\alpha) + (Q_{\alpha\gamma} G_{\gamma\beta} - G_{\alpha\gamma} Q_{\gamma\beta}), \\
\partial_\beta \sigma_{\alpha\beta}^e &= -(c \partial_\alpha \mu + h_\gamma \partial_\alpha p_\gamma + G_{\beta\gamma} \partial_\alpha Q_{\beta\gamma}).
\end{aligned} \tag{3.113}$$

The time invariant derivatives are defined by

$$\begin{aligned}
D_{\alpha\beta} &= \frac{1}{2}(\partial_\alpha v_\beta + \partial_\beta v_\alpha), \quad \Omega_{\alpha\beta} = \frac{1}{2}(\partial_\alpha v_\beta - \partial_\beta v_\alpha), \\
\dot{c} &= \partial_t c + \nabla \cdot (\mathbf{v}c), \quad \dot{P}_\alpha = \partial_t p_\alpha + v_\beta \partial_\beta p_\alpha + \Omega_{\alpha\beta} p_\beta, \\
\overset{\square}{\mathbf{Q}} &= \frac{\partial \mathbf{Q}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{Q} + [\Omega \cdot \mathbf{Q} - \mathbf{Q} \cdot \Omega],
\end{aligned} \tag{3.114}$$

where  $\mathbf{D}$  is the strain rate tensor,  $\Omega$  is the vorticity of the velocity field  $\mathbf{v}$ ,  $P_\alpha$  is the convected co-rotational derivative of  $\mathbf{p}$ ,  $\overset{\square}{\mathbf{Q}}$  is the convected co-rotational derivative of  $\mathbf{Q}$ . If we choose the boundary condition of  $\mathbf{v}$  and  $\mathbf{j}$  so that the boundary integral contribution to the energy dissipation is zero, the time derivative of the total energy reduces to

$$\dot{E}^{total} = - \int_V (\sigma^s, \overset{\square}{\mathbf{Q}}, \dot{\mathbf{P}}, \mathbf{j}) \cdot (\mathbf{D}, \mathbf{G}, \mathbf{h}, -\nabla \mu)^T dV, \tag{3.115}$$

where

$$\begin{array}{ccc}
\text{Flux} & \longleftrightarrow & \text{Force} \\
(\sigma_{\alpha\beta}^s, \overset{\square}{Q}_{\alpha\beta}, \dot{P}_\alpha, j_\alpha) & \longleftrightarrow & (D_{\alpha\beta}, G_{\alpha\beta}, h_\alpha, -\partial_\alpha \mu).
\end{array} \tag{3.116}$$

with this, we propose the following constitutive relation using the generalized Onsager principle:

$$\text{Fluxes} = [\mathbf{M}_{sym} + \mathbf{M}_{anti}] \cdot \text{Force}. \tag{3.117}$$

The total free energy dissipation rate is then given by

$$\frac{d}{dt} \int_V E^{total} d\mathbf{x} = - \int_V T \dot{S} d\mathbf{x} = - \int \text{Force} \cdot \mathbf{M}_{sym} \cdot \text{Forced} \mathbf{x}. \tag{3.118}$$

It is negative provided the symmetric part of the mobility matrix is positive semi-definite. We give some examples to show how the well-known hydrodynamical models are related to the generalized Onsager principle.

**Example 3.4.1** (Binary incompressible viscous fluid mixture model) We have presented some derivations of hydrodynamical models using the variational Onsager principle in the previous section. We now derive it using the constructive Onsager principle approach. We ignore  $\mathbf{p}$  and  $\mathbf{Q}$  and only consider  $c$  as the internal variable. We assume the density is a constant and propose the mobility matrix as follows:

$$\begin{pmatrix} \sigma^v \\ \mathbf{j} \end{pmatrix} = \mathbf{M} \cdot \begin{pmatrix} \mathbf{D} \\ -\nabla\mu \end{pmatrix}, \sigma^a = 0, \nabla \cdot \sigma^e = -c\nabla\mu, \sigma^s = -p\mathbf{I} + \sigma^v, \quad (3.119)$$

where

$$\mathbf{M} = \begin{pmatrix} 2\eta\delta_{\alpha k}\delta_{\beta l} & \\ & \lambda\delta_{\alpha k} \end{pmatrix}, \quad (3.120)$$

$\eta$  is the shear viscosity coefficient and  $\lambda$  is the mobility for the concentration variable  $c$ . The governing system of equations is summarized as follows.

$$\begin{aligned} \frac{\partial c}{\partial t} + \nabla(c\mathbf{v}) + \nabla \cdot \lambda \cdot \nabla\mu &= 0, \\ \nabla \cdot \mathbf{v} &= 0, \end{aligned} \quad (3.121)$$

$$\rho \left( \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla\mathbf{v} \right) = -\nabla p - c\nabla\mu + \nabla \cdot \sigma^s,$$

where  $\rho$  is a constant density. This is known as the Navier-Stokes-Cahn-Hilliard system.

The energy dissipation of the system is given by

$$\frac{d}{dt} E = - \int [2\eta\mathbf{D} : \mathbf{D} + \lambda\|\nabla\mu\|^2] d\mathbf{x}, \quad (3.122)$$

where  $E = \int [\frac{\rho}{2}\|\mathbf{v}\|^2 + f(\phi)] d\mathbf{x}$  is the total free energy. In the Stokes limit, the inertia terms are dropped and the force balance is given by

$$0 = -\nabla p - c\nabla\mu + \nabla \cdot \sigma^s. \quad (3.123)$$

The same energy dissipation rate applies.

**Example 3.4.2** (Quasilinear incompressible model for viscoelastic fluids) We consider only  $\mathbf{Q}$  as the internal variable and ignore the density variation and the polar effect. We apply the Onsager principle to the flux and force pair to arrive at

$$\sigma^s = \sigma^v - p\mathbf{I}, \begin{pmatrix} \sigma^v \\ \square \\ \mathbf{Q} \end{pmatrix} = \mathbf{M} \cdot \begin{pmatrix} \mathbf{D} \\ \mathbf{G} \end{pmatrix}, \quad (3.124)$$

where

$$\mathbf{M}_{sym} = \begin{pmatrix} 2\eta\delta_{\alpha k}\delta_{\beta l} & 0 \\ 0 & \frac{1}{\lambda}\delta_{\alpha k}\delta_{\beta l} \end{pmatrix}, \quad (3.125)$$

$\lambda$  is the relaxation time of the polymer.

$$\mathbf{M}_{anti} = \begin{pmatrix} 0 & -A_1 \\ A_1 & 0 \end{pmatrix}, \quad (3.126)$$

$$A_1 = a[Q_{\alpha k} \delta_{\beta l} + \delta_{\alpha k} Q_{\beta l}],$$

$a \in [-1, 1]$  is a rate parameter. The governing system of equations is given by

$$\begin{aligned} \square \mathbf{Q} - a[\mathbf{D} \cdot \mathbf{Q} + \mathbf{Q} \cdot \mathbf{D}] - \frac{1}{\lambda} \mathbf{G} &= 0, \\ \nabla \cdot (\mathbf{v}) &= 0, \\ \rho \left( \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) &= -\nabla p + \nabla \cdot (\sigma^s - \sigma^e - \sigma^a), \end{aligned} \quad (3.127)$$

where  $\rho$  is a constant density.

The energy dissipation is given by

$$\frac{d}{dt} E = - \int [2\eta \mathbf{D} : \mathbf{D} + \frac{1}{\lambda} \mathbf{G} : \mathbf{G}] d\mathbf{x}, \quad (3.128)$$

where  $E = \int [\frac{\rho}{2} \|\mathbf{v}\|^2 + f(\mathbf{Q})] d\mathbf{x}$  is the total energy. If a quadratic free energy density is given, the Oldroyd B model is recovered [2],

$$f(\mathbf{Q}) = \frac{\gamma}{2} tr(\mathbf{Q}^2), \quad (3.129)$$

where  $\gamma$  is the elastic modulus. If a cubic free energy density is given, the Giesekus model is obtained

$$f(\mathbf{Q}) = \frac{\gamma}{2} tr(\mathbf{Q}^2) + \frac{\gamma_2}{3} tr(\mathbf{Q}^3), \quad (3.130)$$

where  $\gamma_{1,2}$  are elastic moduli. If we choose the free energy density as

$$f(\mathbf{Q}) = \gamma_1 tr(\mathbf{Q}^2) + \gamma_2 (tr(\mathbf{Q}))^2, \quad (3.131)$$

the linear Phan-Thien Tanner model is recovered.

**Example 3.4.3** (Model for nematic liquid crystal solutions) For liquid crystal solutions, we use the polarity vector  $\mathbf{p}$  and the nematic order tensor  $\mathbf{Q}$ . The generalized Onsager principle yields the following constitutive equation.

$$\begin{aligned} (\sigma^s, \square \mathbf{Q}, \dot{\mathbf{P}}, \mathbf{j})^T &= [\mathbf{M}_{sym} + \mathbf{M}_{anti}] \cdot (\mathbf{D}, \mathbf{G}, \mathbf{h}, -\nabla \mu)^T \\ \mathbf{M}_{sym} &= \begin{pmatrix} A_0 & 0 & 0 & 0 \\ 0 & \frac{1}{\gamma_2} \delta_{\alpha k} \delta_{\beta l} & \frac{\chi_1}{2} A & \frac{\chi_2}{2} A \\ 0 & \chi_1 p_\beta \delta_{\alpha k} \delta_{\beta l} & \frac{1}{\gamma_1} \delta_{\alpha k} & \lambda \delta_{\alpha k} \\ 0 & \chi_2 p_\beta \delta_{\alpha k} \delta_{\beta l} & \lambda \delta_{\alpha k} & \gamma \delta_{\alpha k} \end{pmatrix}, \end{aligned} \quad (3.132)$$

where

$$A_0 = 2\eta\delta_{\alpha k}\delta_{\beta l} + (\bar{\eta} - \frac{2}{3}\eta)\delta_{\gamma k}\delta_{\gamma l}\delta_{\alpha\beta} + \alpha_1(Q_{\alpha k}\delta_{\beta l} + \delta_{\alpha k}Q_{\beta l}) \\ + \alpha_2 Q_{kl}Q_{\alpha\beta} + \alpha_3(p_\alpha p_k\delta_{\beta l} + \delta_{\alpha k}p_\beta p_l) + \alpha_4 p_k p_l p_\alpha p_\beta, A = (p_\alpha\delta_{\beta k} + p_\beta\delta_{\alpha k}). \quad (3.133)$$

The coefficient matrix is symmetric and positive definite to ensure energy dissipation.

$$\mathbf{M}_{anti} = \begin{pmatrix} 0 & -A_1 & -A_2 & -A_3 \\ A_1 & 0 & 0 & 0 \\ A'_2 & 0 & 0 & 0 \\ A'_3 & 0 & 0 & 0 \end{pmatrix}, \quad (3.134)$$

$$A_1 = \nu_0 + a[Q_{\alpha k}\delta_{\beta l} + \delta_{\alpha k}Q_{\beta l}] + \nu_3(Q_{kl}Q_{\alpha\beta}) + \theta_1\delta_{kl}\delta_{\alpha\beta},$$

$$A_2 = \frac{\nu_1}{2}(p_\beta\delta_{\alpha k} + p_\alpha\delta_{\beta k}) + \theta_2 p_k\delta_{\alpha\beta}, A'_2 = \nu_1 p_\beta\delta_{\alpha k}\delta_{\beta l} + \theta_2 p_\alpha\delta_{kl},$$

$$A_3 = \frac{\nu_2}{2}(p_\beta\delta_{\alpha k} + p_\alpha\delta_{\beta k}) + \theta_3 p_k\delta_{\alpha\beta}, A'_3 = \nu_2 p_\beta\delta_{\alpha k}\delta_{\beta l} + \theta_3 p_\alpha\delta_{kl}.$$

The coefficient matrix is antisymmetric so that the corresponding part does not contribute to energy dissipation. When the free energy of the liquid crystal is specified, these together with the momentum balance equation and continuity equation  $\nabla \cdot \mathbf{v} = 0$  gives the governing system of equations for the liquid crystal system.

### 3.4.4 Kinetic Theory for Liquid Crystalline Polymer Solutions

The generalized Onsager principle can be applied to mesoscopic modeling. We illustrate it to derive the kinetic equation for liquid crystalline polymers. We model liquid crystalline polymers as rigid rods suspended in a solution. The rod or filament particles are described by their aspect ratio  $a$  and axis of symmetry  $\mathbf{m}$ , with  $\|\mathbf{m}\| = 1$ , and the spatial coordinates  $\mathbf{x}$  of the center of mass. Thus the microstructure configuration space is the sphere  $\mathbf{S}^2$  (for polar rods) or the hemisphere (for apolar rods), and physical space is a domain in  $\mathbf{R}^3$ .

At the kinetic scale, one begins with a microstructure distribution function  $f(\mathbf{x}, \mathbf{m}, t)$  for the rodlike molecule ensemble assuming all rods are identical in size and shape, where  $f(\mathbf{x}, \mathbf{m}, t)d\mathbf{m}$  gives the number of particles with center of mass  $\mathbf{x}$  and orientation  $\mathbf{m}$  within the patch  $d\mathbf{m}$  at time  $t$ . We define the ensemble average in the orientation space  $\mathbf{S}^2 = \{\mathbf{m} \mid \|\mathbf{m}\| = 1\}$  by [2, 4]

$$\langle (\cdot) \rangle = \int_{\mathbf{S}^2} (\cdot) f(\mathbf{m}, \mathbf{x}, t) d\mathbf{m}. \quad (3.135)$$

We denote the zeroth, first, and second moments of  $f$  by  $c = \langle 1 \rangle$ ,  $\mathbf{p} = \langle \mathbf{m} \rangle$ ,  $\mathbf{M} = \langle \mathbf{m}\mathbf{m} \rangle$ , respectively, where  $c$  is the *rod number density*,  $\mathbf{n} = \frac{1}{c}\mathbf{p}$  is the *polarity vector*, and  $\mathbf{Q} = \frac{1}{c}\mathbf{M} - \frac{1}{3}\mathbf{I}$  is the *orientation tensor* [1, 6, 7, 9, 19]. (Note: different authors normalize  $c$  to get a particle, mass, or volume fraction.)

In rodlike molecule systems, individual rods are transported by a superposition of four effects: (1) macroscopic velocity  $\mathbf{u}$  of the fluid mixture, (2) diffusive translational transport in  $\mathbf{x}$ -space due to spatial gradients of the chemical potential, (3) rotational velocity induced by the macroscopic flow and modeled by the Jeffery orbit equation in eqn. (3.140), and (4) diffusive rotational transport in  $\mathbf{m}$ -space due to rotational gradients of the chemical potential. A key remaining ingredient entails the nonlocal interactions of high aspect ratio rods and filaments, which we now discuss.

The nonlocal microstructure interaction potential has a general form,

$$U = \int_V \int_{\mathcal{S}^2} K(\mathbf{m}, \mathbf{m}', \mathbf{x}, \mathbf{y}) f(\mathbf{m}, \mathbf{y}, t) d\mathbf{m}' d\mathbf{y}, \quad (3.136)$$

where  $K(\mathbf{m}, \mathbf{m}', \mathbf{x}, \mathbf{y})$  is an interaction kernel [4, 9, 10, 20] and  $V \in \mathbf{R}^3$  is the physical domain for rods in  $\mathbf{R}^3$ . In practice,  $U$  is typically approximated by a local representation in terms of low moments (shown here up to 2nd moments) and their derivatives via a truncation of the series expansion of the kernel function:

$$U \approx U(c, \nabla c, \mathbf{p}, \nabla \mathbf{p}, \mathbf{M}, \nabla \mathbf{M}). \quad (3.137)$$

The Onsager and Maier-Saupe excluded-volume potentials for liquid crystals are two classical examples, while Marrucci and Greco [10, 20] extended the potential for nematic polymers to incorporate gradient elasticity (the kinetic analog of Frank elasticity).

Combining the nonlocal potential and the entropy of the system, the *free energy* over the material volume  $V$  for the system is given by

$$F[f] = k_B T \int_V \int_{\mathcal{S}^2} \left[ \ln f - f + \frac{U}{2} \right] f d\mathbf{m} d\mathbf{x} = k_B T \int_V \left\langle \left[ \ln f - f + \frac{U}{2} \right] \right\rangle d\mathbf{x}. \quad (3.138)$$

The *chemical potential* is then given by

$$\mu = k_B T \frac{\delta F}{\delta f} = k_B T [\ln f + U]. \quad (3.139)$$

The transport equation for the probability density function  $f$  is given by the *Smoluchowski equation* that couples advection, physical and configurational space diffusion, and rotation by the flow:

$$\frac{\partial f}{\partial t} + \nabla \cdot (\mathbf{u}f) + \mathcal{R} \cdot ((\mathbf{m} \times (\dot{\mathbf{m}}f)) = \nabla \cdot \mathbf{j}_x + \mathcal{R} \cdot \mathbf{j}_r,$$

$$\dot{\mathbf{m}} = \Omega \cdot \mathbf{m} + a [\mathbf{D} \cdot \mathbf{m} - \mathbf{D} : \mathbf{m}\mathbf{m}],$$



where operator  $\mathcal{R} = \mathbf{m} \times \frac{\partial}{\partial \mathbf{m}}$  is the rotational gradient,  $\dot{\mathbf{m}}$  is the Jeffery orbit for the rodlike particle with the aspect ratio  $a$  in Stokes flow,  $\mathbf{D} = \frac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^T)$  is the rate of strain tensor, and  $\Omega = \frac{1}{2}(\nabla \mathbf{u} - \nabla \mathbf{u}^T)$  is the vorticity tensor, respectively.

Next, we turn to the remaining hydrodynamic equations and extra stress contributions arising from the microstructure. We assume the mass and momentum conservation equation given by,

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) &= 0, \\ \rho \left[ \frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right] &= \nabla \cdot (-\Pi \mathbf{I} + \tau) + \mathbf{F}_e, \end{aligned} \quad (3.140)$$

where  $\rho$  is the density of the fluid,  $\Pi$  is the hydrostatic pressure, the extra stress is given by

$$\tau = \tau^{sym} + \tau^{antisym}, \quad (3.141)$$

and  $\mathbf{F}_e$  is the elastic force. We denote the total energy by

$$E = \int_V \left[ \frac{\rho}{2} \|\mathbf{v}\|^2 dx + F[f] \right]. \quad (3.142)$$

Then, the total energy dissipation is calculated as follows

$$\begin{aligned} \frac{dE}{dt} &= \int_V [\mathbf{v} \cdot \nabla \cdot (-\rho \mathbf{v} \mathbf{v} + \tau - \Pi \mathbf{I} + \mathbf{F}_e) + \int_{S^2} \frac{\delta F}{\delta f} \frac{\partial f}{\partial t} d\mathbf{m}] dx \\ &= \int_V [\mathbf{v} \cdot \nabla \cdot (-\rho \mathbf{v} \mathbf{v} + \tau - \Pi \mathbf{I} + \mathbf{F}_e) + \int_{S^2} [-\frac{\delta F}{\delta f} \nabla \cdot (\mathbf{v} f) \\ &\quad + \nabla \cdot \mathbf{j}_x - \mathcal{R} \cdot (f \mathbf{m} \times (\Omega \cdot \mathbf{m} - a \mathbf{D} \cdot \mathbf{m} \mathbf{m}) - \mathbf{j}_m)] d\mathbf{m}] dx \\ &= - \int_V [\mathbf{D} : (\tau^{sym} + \frac{a}{2} (\mathbf{m} \times \mathcal{R} \mathbf{m} + \mathbf{m} \mathbf{m} \times \mathcal{R} \mathbf{m}) - \int_{S^2} (\nabla \mu \cdot \mathbf{j}_x + \mathcal{R} \mu \cdot \mathbf{j}_m) d\mathbf{m}] dx \\ &\quad + \int_{\partial V} \mathbf{n} \cdot [-\frac{\rho}{2} \|\mathbf{v}\|^2 + \tau^{sym} \cdot \mathbf{v} - p \mathbf{v} - \int_{S^2} (\mathbf{j}_x + \mathbf{v} \mu f) d\mathbf{m}] dx. \end{aligned} \quad (3.143)$$

We let

$$\begin{aligned} \mathbf{F}_e &= -\langle \nabla \mu \rangle, \\ \tau^{antisym} &= -\frac{1}{2} \langle \mathbf{m} \times \mathcal{R} \mu \mathbf{m} - \mathbf{m} \mathbf{m} \times \mathcal{R} \mu \rangle. \end{aligned} \quad (3.144)$$

We the generalized Onsager principle to yield the following linear response equation

$$\begin{pmatrix} \tau^{sym} \\ \mathbf{j}_x \\ \mathbf{j}_m \\ r \end{pmatrix} = \begin{bmatrix} C & 0 & 0 \\ 0 & D_s f & 0 \\ 0 & 0 & D_r f \end{bmatrix} \cdot \begin{pmatrix} \mathbf{D} \\ -\nabla \mu \\ -\mathcal{R} \mu \end{pmatrix} \quad (3.145)$$

where  $C$  is a fourth order tensor known as the friction coefficient. If we choose  $C_{ijkl} = 2\eta\delta_{ik}\delta_{jl} + \nu\delta_{ij}\delta_{kl}$ , the coefficient leads to the viscous stress, where  $\eta$  is the shear viscosity and  $\nu$  the volumetric viscosity. We write the elastic body force  $\mathbf{F}_e$  using the *Ericksen stress*  $\tau_e$  defined by  $\nabla \cdot \tau_e = \mathbf{F}_e$ .

We summarize the equations in the model as follows.

$$\begin{aligned}\nabla \cdot \tau_e &= -\langle \nabla \mu \rangle, \\ \tau^{sym} &= \tau_e + 2\eta \mathbf{D} + \nu \text{tr}(\mathbf{D})\mathbf{I} - a/2[\langle \mathbf{m} \times \mathcal{R}\mu \mathbf{m} + \mathbf{m} \mathbf{m} \times \mathcal{R}\mu \rangle], \\ \tau^{antisym} &= -1/2[\langle \mathbf{m} \times \mathcal{R}\mu \mathbf{m} - \mathbf{m} \mathbf{m} \times \mathcal{R}\mu \rangle], \tau^v = 2\eta \mathbf{D} + \nu \text{tr}(\mathbf{D})\mathbf{I}, \quad (3.146) \\ \mathbf{j}_x &= -D_s f \cdot \nabla \mu, \\ \mathbf{j}_m &= -D_r f \cdot \mathcal{R}\mu.\end{aligned}$$

Then, the Smoluchowski equation for  $f$  takes a general, transparent, and conservative form:

$$\partial f / \partial t + (\nabla, \mathcal{R}) \cdot (\mathbf{u}f + -D_s f \cdot \nabla \mu, \mathbf{m} \times \dot{\mathbf{m}}f - D_r f \cdot \mathcal{R}\mu) = 0. \quad (3.147)$$

In this way, we have “derived” a kinetic theory for a liquid crystal solution, which is a two scale model. This generalized Onsager relation lays the foundation for multiscale kinetic theories for complex fluids flows, including both solutions and melts.

### 3.5 Conclusion

We have discussed the Onsager principle for irreversible nonequilibrium processes and the generalized Onsager principle for both reversible and irreversible nonequilibrium processes. The constructive formulation of the Onsager principle applies to general nonequilibrium processes while the Onsager maximum action potential principle only applies to the irreversible processes. We demonstrate by examples that the Onsager principle is an effective modeling tool to develop dynamical theories for any nonequilibrium systems.

### References


1. Baskaran, A., Marchetti, M.: Statistical mechanics and hydrodynamics of bacterial suspensions. Proc. Natl. Acad. Sci. USA **106**(37), 15567–15572 (2009)
2. Bird, B., Armstrong, R.C., Hassager, O.: Dynamics of polymeric liquids. Volume 1: Fluid Mechanics. John Wiley and Sons, New York (1987)

3. de Groot, S.R., Mazur, P.: *Non-Equilibrium Thermodynamics*. Dover Publications, Inc. (1984)
4. Doi, M., Edwards, S.F.: *The Theory of Polymer Dynamics*. Oxford Science Publication (1986)
5. Forest, M.G., Phuworawong, P., Wang, Q., Zhou, R.: Rheological signatures in limit cycle behavior of dilute, active, polar lcps in steady shear. *Phil. Trans. R. Soc. A* **372**: 20130362(1–21) (2014)
6. Forest, M.G., Wang, Q., Zhou, R.: Kinetic theory and simulations of active polar liquid crystalline polymers. *Soft Matters* **9**:5207–5222 (2013)
7. Huang, K.: *Statistical Mechanics*. John Wiley & Sons (2004)
8. Liverpool, T.B., Marchetti, M.C.: Hydrodynamics and rheology of active polar filaments. In: Lanz, P (ed.), *Cell Motility*, pp. 177–206. Springer, New York (2008)
9. Marrucci, G., Greco, F.: Flow behavior of liquid crystalline polymers. *Adv. Chem. Phys.* **86**, 331–404 (1993)
10. Masao, D.: *Soft Matter Physics*. Oxford University Press (2013)
11. Mazenko, G.F.: *Nonequilibrium Statistical Mechanics*. Wiley-VCH (2006)
12. Onsager, L., Machlup, S.: Fluctuations and irreversible processes. II. systems with kinetic energy. *Phys. Rev.* **91**:1512–1515 (1953)
13. Onsager, L.: Reciprocal relations in irreversible processes. i. *Phys. Rev.* **37**:405–426 (1931)
14. Onsager, L.: Reciprocal relations in irreversible processes II. *Phys. Rev.* **38**, 2265–2279 (1931)
15. Onsager, L., Machlup, S.: Fluctuations and irreversible processes. *Phys. Rev.* **91**, 1505–1512 (1953)
16. Pathria, R.K.: *Statistical Mechanics*, 2nd ed. Butterwoths and Heinemann (2001)
17. Plischke, M., Bergersen, B.: *Equilibrium Statistical Physics*. World Scientific (2006)
18. Qian, Tiezheng., Wang, Xiao-Ping, Sheng, Ping: A variational approach to moving contact line hydrodynamics. *J. Fluid Mech.* **564**, 333–360 (2006)
19. Saintillan, D., Shelley, M.J.: Active suspensions and their nonlinear models. *Comptes Rendus Physique* **14**, 497 (2013)
20. Wang, Q.: A hydrodynamic theory for solutions of nonhomogeneous nematic liquid crystalline polymers of different configuration. *J. Chem. Phys.* **116**(20), 9120–9136 (2002)
21. Yang, X., Li, J., Gregory Forest, M., Wang, Q.: Hydrodynamic theories for flows of active liquid crystals and the generalized on sager principle. *Entropy* **18**:202 (2016)

# Chapter 4

## An Introduction to Emergence Dynamics in Complex Systems



Zhigang Zheng 

**Abstract** Emergence is one of the most essential features of complex systems. This property implies new collective behaviors due to the interaction and self-organization among elements in the system, which cannot be produced by a single unit. It is our task in this Chapter to extensively discuss the basic principle, the paradigm, and the methods of emergence in complex systems based on nonlinear dynamics and statistical physics. We develop the foundation and treatment of emergent processes of complex systems, and then exhibit the emergence dynamics by studying two typical phenomena. The first example is the emergence of collective sustained oscillation in networks of excitable elements and gene regulatory networks. We show the significance of network topology in leading to the collective oscillation. By using the dominant phase-advanced driving method and the function-weight approach, fundamental topologies responsible for generating sustained oscillations such as Winfree loops and motifs are revealed, and the oscillation core and the propagating paths are identified. In this case, the topology reduction is the key procedure in accomplishing the dimension-reduction description of a complex system. In the presence of multiple periodic motions, different rhythmic dynamics will compete and cooperate and eventually make coherent or synchronous motion. Microdynamics indicates a dimension reduction at the onset of synchronization. We will introduce statistical methods to explore the synchronization of complex systems as a non-equilibrium transition. We will give a detailed discussion of the Kuramoto self-consistency approach and the Ott-Antonsen ansatz. The synchronization dynamics of a star-networked coupled oscillators and give the analytical description of the transitions among various ordered macrostates. Finally, we summarize the paradigms of studies of the emergence and complex systems.

**Keywords** Emergence · Order parameter · Slaving principle · Self-sustained oscillation · Winfree loop · Synchronization · Complex networks

---

Z. Zheng (✉)

Institute of Systems Science and College of Information Science and Engineering,  
Huaqiao University, Xiamen 361021, China  
e-mail: [zgzheng@hqu.edu.cn](mailto:zgzheng@hqu.edu.cn)

## 4.1 Introduction

The core mission of physics is to understand basic laws of everything in the universe. The Chinese vocabulary the English word “UNIVERSE” is “宇宙”, which is composed of two characters “宇” and “宙” with distinctly different meanings. The word “宇” means the space, and “宙” refers to the time. The ancient Chinese philosopher Shi Jiao (尸佼, or 尸子, Shi Zi) in the Warring States period (475–221 B.C.), wrote that “四方上下曰宇, 往古来今曰宙” in his book “Shi Zi”. Similar expressions also appeared in other books such as “Wen Zi: the Nature”, “Zhuang Zi. Geng Sang Chu”, “Huai Nai Zi”, and so on [1]. These interesting citations indicate that the mission of physicists is a deep understanding of basic laws of space and time, or say, the common and fundamental laws of variations embedded in different systems.

A basic paradigm of natural science developed throughout the past many decades is the principle of reductionism. The essence of reductionism is that a system is composed of many elements, with each element can be well understood and physically described. As long as the laws governing the elements are clear, it is expected that the properties of the system can be well understood and reconstructed. This belief had been successfully undertaken in the 18–19 centuries, where the structures of matters, ranging from molecules, atoms, protons, neutrons, electrons to quarks, were successfully revealed [2].

The reductionism encountered its crisis from the first half of the twentieth century. The development of thermodynamics and statistical physics, especially the findings of various phase transitions such as superconductivity and superfluidity, indicated that these behaviors result from the collective macroscopic effect of molecules and atoms, which do not occur at the atom/microscopic level. Physicists working on statistical physics interpreted the occurrence of phase transition as a symmetry breaking, which can be extensively found in antiferromagnets, ferroelectrics, liquid crystals and condensed matters in many other states. Philip Andersen connected the later findings of quasiparticles, e.g. phonons, implies collective excitations and modes organized by atoms with interactions [3]. Collective behaviors that are formed by interacted units in a system, e.g. transitions among different phases, are termed as the *emergence*.

The emergence property of many-body systems implies the collapse and failure of reductionism. Andersen asserted:

The ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the universe..... The behavior of large and complex aggregates of elementary particles, it turns out, is not to be understood in terms of a simple extrapolation of the properties of a few particles. Instead, at each level of complexity entirely new properties appear, and the understanding of the new behaviors requires research which I think is as fundamental in its nature as any other.

Nowadays, emergent behaviors have been extensively found in living activities of neuron systems, brains and the formation of proteins, DNA, and genes [4]. These phenomena imply that even if one knows everything about an element, the behaviors

of a system composed of these elements still cannot be simply predicted based on individual properties. All these are called **complex systems**.

One can list some common properties of complex systems. When looked at in detail, such as cooperation or self-organization, emergence, and adaption [5]. The process of organized behavior arising without an internal or external controller or leader in a system is called the **self-organization**. Since simple rules produce complex behavior in hard-to-predict ways, the macroscopic behavior of such systems is sometimes called the **emergence**. A complex system is then often defined as a *system that exhibits nontrivial emergent and self-organizing behaviors*. The central question of the sciences of complexity is how these emergent self-organized behaviors come about [6].

Emergence is one of the most essential features of complex systems. In this Chapter, we will discuss extensively the basic principle, the paradigm, and the methods of emergence in complex systems based on nonlinear dynamics and statistical physics. We develop the foundation and treatment of emergent processes of complex systems, and then exhibit the emergence dynamics by studying two typical phenomena.

The first example is the emergence of collective sustained oscillation in networks of excitable elements and gene regulatory networks. We show the significance of network topology in leading to the collective oscillation. By using the dominant phase-advanced driving method and the function-weight approach, fundamental topologies responsible for generating sustained oscillations such as Winfree loops and motifs are revealed, and the oscillation core and the propagating paths are identified. In this case, the topology reduction is the key procedure in accomplishing the dimension-reduction description of a complex system.

In the presence of multiple periodic motions, different rhythmic dynamics will compete and cooperate and eventually make coherent or synchronous motion. Micro-dynamics indicates a dimension reduction at the onset of synchronization. We will introduce statistical methods to explore the synchronization of complex systems as a non-equilibrium transition. We will give a detailed discussion of the Kuramoto self-consistency approach and the Ott-Antonsen ansatz. The synchronization dynamics of a star-networked coupled oscillators gives the analytical description of the transitions among various ordered macrostates.

We will summarize the paradigms of studies of the emergence and complex systems based on the above discussions.

## 4.2 Emergence: Research Paradigms

Emergence implies the self-organized behavior in a complex system, which occurs under the physically non-equilibrium condition. This collective feature comes from the cooperation of elements through coupling, and it cannot be observed at the

microscopic level. To reveal the emergent dynamics at the macroscopic level, scientists have proposed various theories from microscopic to statistical and macroscopic viewpoints.

### 4.2.1 Entropy Analysis and Dissipative Structure

Let us first discuss the possibility of self-organization in non-equilibrium systems from the viewpoint of thermodynamics and statistical physics. This implicitly requests an open system that can exchange matters, energy, and information with its environment, as shown in Fig. 4.1. We focus on the entropy change  $dS$  in a process of an open system. One may decompose the total entropy production  $dS$  into the sum of two contributions:

$$dS = d_iS + d_eS, \quad (4.1)$$

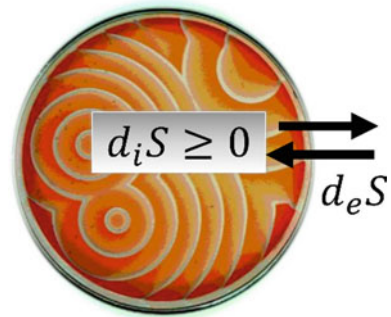
where  $d_iS$  is the entropy production due to the irreversible process inside the system, and  $d_eS$  is the entropy flux due to the exchanges with the environment. The second thermodynamic law implies that

$$d_iS \geq 0, \quad (4.2)$$

where  $d_iS = 0$  denotes the thermal equilibrium state. If the system is isolated,  $d_eS = 0$ , one has  $dS = d_iS > 0$ . In the presence of exchanges with the environment,  $d_eS \neq 0$ . When this open system reaches the steady state, i.e. the total entropy change  $dS = 0$ . This leads to

$$dS = -d_eS < 0. \quad (4.3)$$

**Fig. 4.1** A schematic entropy process of an open system exchanging with the environment, which leads to the emergence of non-equilibrium structure



This means that if there exists a sufficient amount of negative entropy flow, the system can be expected to maintain an ordered configuration. Prigogine and colleagues thus claimed that “*Nonequilibrium may be a source of order*”, which forms the base of the dissipative structure theory [7].

The emergence of dissipative structure depends on the degree of deviation from the equilibrium state of the system. In the small-deviation regime, the system still keeps its thermodynamic property, and the minimum entropy principle applies. In the linear regime, numerous theories such as linear response theory and dissipation-fluctuation theorem have been proposed. As the system is driven so far from the equilibrium state that the thermodynamic branch becomes unstable, structural branches may emerge and replace the thermodynamic branch [7, 8]. This can be mathematically described in terms of dynamical system theory [9, 10].

Denote the macrostate of a complex system as  $\vec{u}(t) = (u_1, u_2, \dots, u_n)$ , the evolution of the state can be described as

$$d\vec{u}/dt = \vec{f}(\vec{u}, \boldsymbol{\epsilon}), \quad (4.4)$$

where  $\vec{f} = (f_1, f_2, \dots, f_n)$  is the nonlinear function vector, and  $\boldsymbol{\epsilon}$  are a group of control parameters. Equation (4.4) is usually a group of coupled nonlinear equations and can be extensively discussed by using theories of dynamic systems, and the stability of possible states and bifurcations have been exhaustively studied in the past decades. Readers can refer any textbook on nonlinear dynamics and chaos to gain a detailed understanding [11, 12].

Considering the spatial effect, i.e.  $u = u(\mathbf{r}, t)$ . The simplest spatial effect in physics is the diffusion process, which is given by Fick’s law as the proportional relation between the flux and the gradient of matter condensation in space:

$$\mathbf{J} = -D\nabla u, \quad (4.5)$$

where  $D$  is the diffusion coefficient. Therefore in this case the governing equation of motion can be written as

$$\partial u / \partial t = f(u, \boldsymbol{\epsilon}) + D\nabla^2 u, \quad (4.6)$$

where  $D$  denotes the diffusion coefficient. Equation (4.6) is called the *reaction–diffusion equation*. This equation and related mechanism were first proposed by Alan Turing in 1952 as a possible source of biological organism [13]. The reaction term  $f(u, \boldsymbol{\epsilon})$  gives the local dynamics, which is the source of spatial inhomogeneity. The diffusion term tends to erase the spatial differences of the state  $u$ , thus it is the source of spatial homogeneity. Both mechanisms appears at the right hand side of Eq. (4.6) and compete, resulting in a self-organized state. Equation (4.6) and related dynamical systems have been extensively explored in the past a few years, and rich



spatiotemporal patterns and dynamics have been revealed. Readers can reach related reviews and monographs to get more information [14–16].

### 4.2.2 *Slaving Principles and the Emergence of Order Parameters*

The study of dissipative structure is, in fact, largely based on the dynamics of macrostate variables. However, it is very important to appropriately select these macroscopic variables. These state variables are required to reveal the emergence of dissipative structures, hence they should act as *order parameters* similar to studies of those in phase transitions. The concept of order parameter was first introduced in statistical physics and thermodynamics to describe the emergence of order and the transitions of a thermodynamic system among different macroscopic phases [17–19]. It had been naturally extended to non-equilibrium situations to reveal the order out of equilibrium.

Haken and his collaborators proposed the synergetic theory and focused on the conditions, features and evolution laws of the self organization in a complex system with a large number of degrees of freedom under the drive of external parameters [20, 21], and the interaction between subsystems to form spatial, temporal or functional ordered structures on a macroscopic scale. The core of synergetics is the slaving principle, which reveals how order parameters emerge from a large number of degrees of freedom through competitions and collaborations. As the system approaches the critical point, only a small number of modes/variables with a slow relaxation dominate the macroscopic behavior of the system and characterize the degree of order (called the order parameters) of the system. A large number of fast-changing modes are governed by the order parameters and can be eliminated adiabatically. Thus we can establish the basic equation of the order parameters. The low-dimensional evolution equation of order parameters can thus be used to study the emergence of various non-equilibrium states, their stability and bifurcations/transitions.

To clarify the emergence of order parameters, let us first take a simple two-dimensional nonlinear dynamical system as an example. Suppose the following nonlinear differential equations with two variables ( $u(t)$ ,  $s(t)$ ):

$$\dot{u} = \alpha u - us, \quad (4.7a)$$

$$\dot{s} = -\beta s + u^2. \quad (4.7b)$$

where the linear coefficients are  $\alpha$ ,  $\beta$ , and  $\beta > 0$ . Let us set  $\alpha$  as the modulated parameter. When  $\alpha < 0$ , the stationary solution is  $(u, s) = (0, 0)$ . By changing the parameter  $\alpha$  to slightly larger than 0, i.e.  $0 < \alpha \ll 1$ , the solution  $(u, s) = (0, 0)$  becomes unstable. The new solution

$$s = \alpha, u = \sqrt{\alpha\beta} \quad (4.8)$$

emerges and keeps stable. However, because the new stable solution (4.8) is still near (0,0), (4.7b) can be solved by integrating (4.7b) approximately as

$$\begin{aligned} s(t) &= \int_0^t e^{-\beta(t-\tau)} u^2(\tau) d\tau \\ &= \frac{1}{\beta} u^2(t) - \frac{2}{\beta} \int_0^t e^{-\beta(t-\tau)} u(\tau) \dot{u}(\tau) d\tau. \end{aligned} \quad (4.9)$$

The first expression is the integral form, and the second expression can be obtained by using the partial integral. Considering  $0 < \alpha \ll 1$ , and the variables  $u$ ,  $s$ ,  $\dot{u}$ ,  $\dot{s}$  are also small but with different order. Using the simple scaling analysis, one can get

$$u \sim \sqrt{\alpha}, s \sim \alpha, \dot{u} \sim \alpha^{3/2}, \dot{s} \sim u\dot{u} \sim \alpha^2. \quad (4.10)$$

Therefore  $|\dot{u}| \ll |u|$  when  $\alpha \ll 1$ , the second term in the second expression of (4.9) is a high-order term and can be neglected. Thus one obtains the following approximated equation:

$$s(t) \approx \frac{1}{\beta} u^2(t). \quad (4.11)$$

By comparing (11a) and (4.7b), one can easily find that the result of (11a) is equivalent to setting

$$\dot{s} = 0 \quad (4.12)$$

in (4.7b), and one can get  $s(t) = u^2(t)/\beta$ . Substituting it to (4.7a) one obtains

$$\dot{u} = \alpha u - \frac{1}{\beta} u^3. \quad (4.13)$$

This is a one-dimensional dynamical equation and can be easily solved.

The proposition (4.12) is called the **adiabatic elimination principle**, which is a commonly used approximation method adopted in applied mathematics. This approximation has a profound physical meaning. Eq. (4.12) indicates that, under this condition, the variable  $u(t)$  is a slow-varying and linearly-unstable mode called the **slow variable**, while  $s(t)$  is a fast-changing and linearly-stable mode called the **fast variable**. The real essence of (4.12) is that, near the onset of the critical point, the fast mode  $s(t)$  can be so fast that it can always keep up with the change of the slow mode  $u(t)$ , and the fast variable can be considered as the function of the slow variable, as

shown in (4.11). In other words, the slow mode dominates the evolution of the system and of course, the fast mode can be reduced in terms of adiabatic elimination, leaving only the equation of the slow mode. Therefore, the slow mode will determine the dynamical tendency of the system (7) in the vicinity of the critical point and can be identified as the *order parameter*. The emergence of order parameters in a complex system is the central point of the **slaving principle**, which was proposed by Hermann Haken [21].

The above discussion exhibits a typical competition of two modes in a two-dimensional dynamical system. The insightful thought embedded in the slaving principle can be naturally extended to complex systems with a large number of competing modes. Suppose an  $n$ -dimensional dynamical system  $\dot{\mathbf{x}} = \mathbf{F}(\alpha, \mathbf{x})$ , where  $\alpha$  is the controlling parameter. The equations of motion can be written in the following canonical form near the critical point  $\mathbf{x} = 0$ :

$$\dot{\mathbf{x}} = \mathbf{A}(\alpha)\mathbf{x} + \mathbf{B}(\mathbf{x}, \alpha), \quad (4.14)$$

where  $\mathbf{x}$  is an  $n$ -dimensional state vector  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ ,  $\mathbf{A} = \mathbf{A}(\alpha)$  is an  $n \times n$  Jacobian matrix, and  $\mathbf{B}(\mathbf{x}, \alpha)$  is an  $n$ -dimensional nonlinear function vector of  $\mathbf{x}$ . The eigenvalues of the matrix  $\mathbf{A}$  are  $\{\lambda_i(\alpha)\}$ , which are aligned as the descending order according to their real parts, i.e.

$$\text{Re}\lambda_1 \geq \text{Re}\lambda_2 \geq \dots \geq \text{Re}\lambda_n.$$

Assume that  $\mathbf{x} = 0$  is a stable solution of (4.14) in a certain parameter regime of  $\alpha$ , i.e. the real parts of all eigenvalues are negative,

$$\text{Re}\lambda_i(\alpha) < 0, i = 1, \dots, n. \quad (4.15)$$

By modulating the parameter  $\alpha$  to a critical point, say,  $\alpha > \alpha_c$ , when  $\text{Re} \lambda_1$  changes from negative to positive, i.e.

$$0 < \text{Re}\lambda_1 \ll 1, \quad (4.16)$$

and other eigenvalues  $\{\text{Re}\lambda_2, \text{Re}\lambda_3, \dots, \text{Re}\lambda_n\}$  remain negative. In this case the solution  $\mathbf{x} = 0$  becomes unstable. By introducing the linear transformation matrix  $\mathbf{T}$  of the Jacobian  $\mathbf{A}$  as

$$\tilde{\mathbf{A}} = \mathbf{T}^{-1}\mathbf{A}\mathbf{T} \quad (4.17)$$

so that the new matrix is diagonalized as.

$$\tilde{A} = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \lambda_n \end{pmatrix} \quad (4.18)$$

To distinguish the eigenvalue  $\lambda_1$  from other eigenvalues, we relabel these eigenvalues as.

$$\lambda_u = \lambda_1, \lambda_s^1 = \lambda_2, \lambda_s^2 = \lambda_3, \dots, \lambda_s^{(n-1)} = \lambda_n,$$

where  $\text{Re}\lambda_{s_i} < 0$ ,  $i = 1, 2, \dots, n-1$ . Then the corresponding state vector  $\mathbf{x}$  can be transformed to.

$$(u, \mathbf{s})^T = \mathbf{T}\mathbf{x}. \quad (4.19)$$

Equations (4.14) can be rewritten as

$$\dot{u} = \lambda_u u + \tilde{B}_u(u, \mathbf{s}), \quad (4.20a)$$

$$\dot{\mathbf{s}} = -\lambda_s \mathbf{s} + \tilde{B}_s(u, \mathbf{s}), \quad (4.20b)$$

where  $\mathbf{s}$  is an  $(n-1)$ -dimensional vector  $\mathbf{s} = (s_1, s_2, \dots, s_{n-1})^T$ .

By using the above procedure, one successfully separates the slow mode  $u(t)$  from all the variables in terms of the transformation (4.20a), and the remaining variables  $\{s_i(t), i = 1, 2, \dots, n-1\}$  are fast variables that satisfy Eq. (4.20b). One can apply the adiabatic elimination to (4.20b) based on the same reason as

$$\dot{\mathbf{s}} = 0. \quad (4.21)$$

This leads to the following  $n-1$  equations:

$$\lambda_s \mathbf{s} = -\tilde{B}_s(u, \mathbf{s}). \quad (4.22)$$

Fast variables  $\mathbf{s}$  can be analytically solved from the  $n-1$  equations of (4.22) as the function of the slow variable  $u$  in the form  $\mathbf{s} = \mathbf{s}(u)$ . Then by inserting  $\mathbf{s}(u)$  into Eq. (4.20a), one obtains

$$\dot{u} = \lambda_u u + \tilde{B}_u(u, \mathbf{s}(u)). \quad (4.23)$$

This is the one-dimensional dynamical equation of the order parameter  $u$ , which can be easier to analyze.

When there exist degenerations for the first  $m > 1$  eigenvalues  $\{\lambda_1, \lambda_2, \dots, \lambda_m\}$ , which means that they have the same real parts:

$$\operatorname{Re}\lambda_1(\alpha) = \operatorname{Re}\lambda_2(\alpha) = \dots = \operatorname{Re}\lambda_m(\alpha), m < n, \quad (4.24)$$

these  $m$  modes may lose their stability simultaneously at the critical point and all are slow modes, and their eigenvalues  $\lambda_u = (\lambda_1, \lambda_2, \dots, \lambda_m)$ . In this case  $u$  and  $\tilde{B}_u$  in (4.23) should be replaced by vectors. Therefore

$$\mathbf{u} = (u_1, u_2, \dots, u_m)^T, \quad (4.25a)$$

$$\mathbf{s} = (s_1, s_2, \dots, s_{n-m})^T, \quad (4.25b)$$

$$\tilde{\mathbf{B}}_u = (\tilde{B}_u^1, \tilde{B}_u^2, \dots, \tilde{B}_u^m)^T, \quad (4.25c)$$

and the equations of motion are rewritten as

$$\dot{\mathbf{s}} = -\lambda_s \mathbf{s} + \tilde{\mathbf{B}}_s(\mathbf{u}, \mathbf{s}), \dot{\mathbf{u}} = \lambda_u \mathbf{u} + \tilde{\mathbf{B}}_u(\mathbf{u}, \mathbf{s}). \quad (4.26)$$

$\mathbf{s}(\mathbf{u})$  can be obtained in terms of the adiabatic elimination (4.20b). Eq. (4.23) becomes  $m$ -dimensional equations of motion by inserting the formula  $\mathbf{s}(\mathbf{u})$  into (4.20a):

$$\dot{\mathbf{u}} = \lambda_u \mathbf{u} + \tilde{\mathbf{B}}_u(\mathbf{u}, \mathbf{s}(\mathbf{u})). \quad (4.27)$$

These are the equations of motion of the order parameters  $\mathbf{u} = (u_1, u_2, \dots, u_m)^T$ . By comparing Eq. (4.27) with Eq. (4.23), one finds that these two equations have the same form. However, they are essentially different. In Eq. (4.26), the  $\mathbf{s}$  variables are functions of time  $t$ , while in Eq. (4.27),  $\mathbf{s}$  are functions of the slow variables  $\mathbf{u}$ , and the degrees of freedom of (4.27) is considerably less than that of (4.26). In practice, only a very small portion of modes may lose their stability at a critical point, therefore one can consider the procedure from Eqs. (4.20) to (4.23) and (4.27) as a reduction from high-dimensional to low-dimensional dynamics governed by only a few order parameters. This obviously is a great dynamical simplification, which is an important contribution of the slaving principle at the critical point.

The slaving principle is closely related to the central manifold theorem in topological geometry. The center-manifold theorem is a commonly used method of dimensionality reduction, which is suitable for studying autonomous dynamical systems. The center-manifold method uses the characteristic of the tangent manifold and the corresponding subspace to find out the equation of the system on the center manifold. For high-dimensional dynamical systems, it is difficult to study the dynamical system directly through the traditional bifurcation behavior. In order to better grasp the nature of the problems to be studied, the central-manifold theorem is generally adopted to reduce the system to lower dimensional equations.

### 4.2.3 *Networks: Topology and Dynamics*

The blossom of network science is absolutely a great milestone in the exploration of complexity. Early studies of complex networks started from the graph theory in mathematics. One can infer from the early work of Euler on the Konigsburg bridge problem. Graph theory proposed a number of useful concepts and laws in analyzing the sets composed of vertices and edges. Although the studies of networks can be traced back to the development of graph theory in mathematics, the scope of today's network science is an interdisciplinary field covering extensive subjects from physics, chemistry, biology, economy, and even social science. The topics of network science thus are explosive with the development of science and technology, the contributions from various subjects vividly enrich the network science.

The important mission of network science is to give a common understanding of various complex systems from the viewpoint of topology. Behind this mission the important issue relates to the reduction of complexity based on network theory and dynamics. Nowadays we can easily get a knowledge and big data via various measure techniques from complex systems. How to gain useful information from these data is essentially a process of reduction, i.e. to get the truth by getting rid of redundant information. To perform such an effective reduction of a complex system from its microdynamics, the very important starting point is to identify an appropriate microscopic description. Two ingredients are indispensable when one studies the microdynamics, i.e. unit dynamics and the coupling patterns of units in a system. Now let me discuss these two points.

Different systems are composed of units with different properties. This is the fundamental viewpoint of reductionism. For example, a drop of water is in fact composed of  $\sim 10^{23}$   $\text{H}_2\text{O}$  molecules, the human brain is composed of  $\sim 10^{11}$  neurons, and a heart tissue is composed of  $\sim 10^{10}$  cardiac cells. Apparently, these units work with different mechanisms and are described by distinct dynamics. It has been a central topic in exploring the mechanism of these units. For example, a single neuron or a cardiac tissue works in an accumulating-firing manner and can be modeled by a mimic nonlinear electronic circuit. This dynamical feature is described by a number of excitable models, e.g. the Hodgkin-Huxley model, the Fitzhugh-Nagumo model, and so on.

The second ingredient is the modelling of interactions among units. At the particle level, the forces among quarks, elementary particles, atoms, molecules are quite different. It is an important task for physicists to explore these interactions. The interactions out of physics are also system dependent. For example, the relations of individuals in a society are complicated, depending strongly on the type of information that is exchanged between two individuals.

It becomes astonishing when there are too many individuals in a system, when the forms of unit dynamics and coupling functions sometimes matter, while in many cases they are not essential. A typical situation happens when the coupling topology is more important than the specific form of coupling function. The behaviors of emergence form when the macroscopic behaviors appear for a complex system while they do not

happen at the level of each unit. Recent studies of the so-called complex networks stimulated by milestone works on small-world [22] and scale-free networks [23] provide a powerful platform in studying complex systems. For example, it is a significant topic to explore network properties of biological gene, DNA, metabolic, and neural networks, social and ecological networks, WWW and internet networks, and so on. Another important topic is the study of dynamical processes on networks, such as synchronization, propagation processes, and network growth. Interested readers may refer related monographs and reviews for more knowledge of network science [24–30].

## 4.3 Emergence of Rhythms

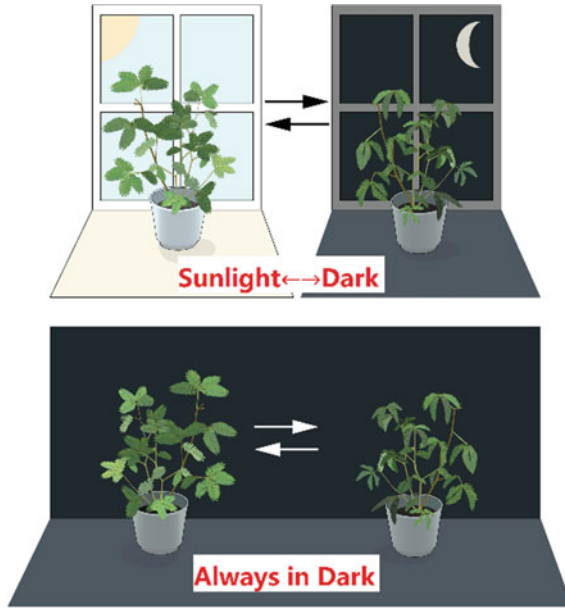
### 4.3.1 *Biological Rhythms: An Introduction*

Biological rhythm is an old question and can be found ubiquitously in various living systems [31]. The 2017 Nobel Prize in Physiology or Medicine was awarded to J. C. Hall, M. Rosbash and M.W. Young for their discoveries of molecular mechanisms that control circadian rhythms [32, 33]. Circadian rhythms are driven by an internal biological clock that anticipates day-night cycles to optimize the physiology and behavior of organisms. The exploration of the emergence of circadian rhythms from the microscopic level (e.g. molecular or genetic levels) aroused a new era of studies on biological oscillations [34] (Fig. 4.2).

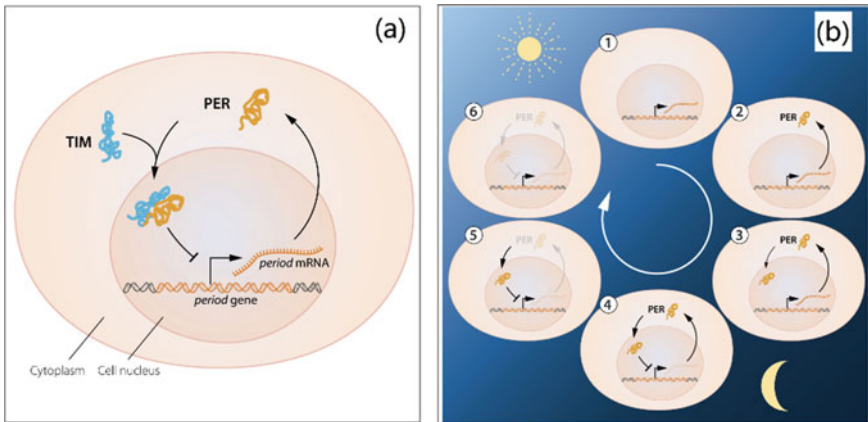
Observations that organisms adapt their physiology and behavior to the time of the day in a circadian fashion have been recorded for a long time. The very early observations of leaf and flower movements in plants, for example, the leaves of mimosa plants close at night and open during the day presented interesting biological clocks. In 1729, the French scientist de Mairan observed that the leaves of a mimosa plant in the dark could still open and close rhythmically at the appropriate time of the day, implying an endogenous origin of the daily rhythm rather than external stimulus [35].

The genetic mechanism responsible for the emergence of circadian rhythms was first explored by S. Benzer and R. Konopka from the 1960s. In 1971, their pioneering work identified mutants of the fruit fly *Drosophila* that displayed alterations in the normal 24-h cycle of pupal eclosion and locomotor activity, which was named as period (PER) [36]. Later on, Hall and Rosbash at Brandeis University [37] and Young at Rockefeller University [38] isolated and molecularly characterized the period gene (Fig. 4.3).

Further studies by Young, Hardin, Hall, Rosbash and Takahashi revealed that the molecular mechanism for the circadian clock relies not on a single gene but on the so-called transcription-translation feedback loop (TTFL), i.e. the transcription of period and its partner gene timeless (TIM) are repressed by the PER and TIM proteins, generating a self-sustained oscillation [39–42]. These explorations led to



**Fig. 4.2** The observation of de Mairan on the rhythmic daily opening-close in the sun and in the dark. **a** in the normal sunlight-dark environment. **b** always in dark., The environment-independent open-closing implies an endogenous origin of the daily rhythm. (Adapted from [33])



**Fig. 4.3** **a** A simplified illustration of the biochemical process of the circadian clock. **b** The transcription-translation feedback loop (TTFL), i.e. the transcription of period and its partner gene timeless (TIM) are repressed by the PER and TIM proteins, generating a self-sustained oscillation. (Adapted from [33])



numerous potential studies that revealed a series of interlocked TTFL's together with a complex network of reactions. These involve regulated protein phosphorylation and degradation of TTFL components, protein complex assembly, nuclear translocation and other post-translational modifications, generating oscillations with a period of approximately 24 h. Circadian oscillators within individual cells respond differently to entraining signals and control various physiological outputs, such as sleep patterns, body temperature, hormone release, blood pressure, and metabolism. The seminal discoveries by Hall, Rosbash and Young have revealed a crucial physiological mechanism explaining circadian adaptation, with important implications for human health and disease.

Essentially, biological rhythms ubiquitously existing in various biological systems are physically oscillatory behaviors, which are indications of temporally periodic dynamics of nonlinear systems. On the other hand, rhythmic phenomena can be extensively observed in various situations, ranging from physics, and chemistry to biology [6, 43]. Therefore, these oscillatory behaviors with completely different backgrounds can be universally studied in the framework of nonlinear dynamics.

#### 4.3.1.1 Self-sustained Oscillation in Simple Nonlinear Systems

*Self-sustained oscillation*, also called the *limit cycle*, which is defined as a typical time-periodic nonlinear behavior, has received much attention throughout the past century [44]. The essential mechanism of sustained oscillation is the existence of feedback in nonlinear systems, but the source of feedback depends strongly on different situations.

First, *the competition-balance mechanism between the positive feedback and the negative feedback can maintain a stable oscillation*. In mechanical systems, for example, the conventional Van der Pol oscillator, the adjustable damping provides the key mechanism for the energy compensation to sustain a stable limit cycle. The damping becomes positive to dissipate energy when the amplitude is large and becomes negative to consume energy when the amplitude is small. Stability analysis can well present the dynamical mechanism of the limit-cycle motion. There are a lot of historic examples and models in describing these typical oscillatory dynamics.

Let us study a simple nonlinear system with a limit cycle oscillation by adopting the following complex dynamical equation:

$$\dot{z} = (\lambda + i\omega)z - bz|z|^2, \quad (4.28)$$

where  $z(t) = x(t) + iy(t)$  is a complex order parameter. We use “i” to denote the imaginary index throughout this chapter. Physically the order parameter  $z(t)$  can be obtained via the reduction procedure by eliminating the fast modes. Equation (4.28) presents a two-dimensional dynamics in real space, and the possible time-dependent solution of this equation is the limit cycle.

To obtain an analytical solution, it is convenient to introduce the polar coordinates.

$$z(t) = r(t)e^{i\theta(t)},$$

and Eq. (4.28) can be decomposed into the following two-dimensional equations of the phase and the amplitude that are uncoupled to each other:

$$\dot{\theta}(t) = \omega, \quad (4.29a)$$

$$\dot{r}(t) = \lambda r - br^3. \quad (4.29b)$$

The phase Eq. (4.29a) indicates that the phase evolves uniformly with the phase velocity  $\omega$ . The amplitude Eq. (4.29b) has two stationary solutions for  $\lambda > 0$ : the unstable solution  $r = 0$  and the stable solution  $r_0 = \sqrt{\lambda/b}$ . By considering the dynamics of both the phase and the amplitude, the latter one represents the periodic solution of Eq. (4.28) as.

$$z(t \rightarrow \infty) = r_0 e^{i(\omega t + \theta_0)}.$$

This is a stable and attractive limit-cycle, and the relaxation process from an arbitrary initial state can lead to this solution. The stability of this sustained oscillator is the result of the competition between the positive feedback term  $\lambda r$  and the negative feedback term  $-br^3$ .

The second source of the feedback mechanism comes from the collaboration of units in a complex system, which is our focus here. In the following discussions, we will study the collective oscillation of a population of interacting non-oscillatory units. Because each unit in the system does not exhibit oscillatory behavior, the feedback mechanism of sustained oscillations should come from the collaborative feedback of units.

The self-sustained oscillation in complex systems consisting of a large number of units is a typical emergence that results from more complicated competitions and self organizations, and the mechanism of this collective oscillation was of great interest in recent years [45–47]. It is thus very interesting and important to explore the mechanism of oscillatory dynamics when these units interact with each other and study how a number of non-oscillatory nodes organize themselves to emerge a collective oscillatory phenomenon [48].

### 4.3.1.2 Self-sustained Oscillations in Complex Systems

The topic of self-sustained oscillations in complex systems was largely motivated by an extensive study and a strong background of system biology. Non-oscillatory systems exist ubiquitously in biological systems, e.g. the gene segments and neurons. People have extensively studied such common behaviors in nature, such

as oscillations in gene-regulatory networks [46, 49–53], neural networks and brains [54–59].

The exploration of the key determinants of collective oscillations is a great challenge. It is not an easy task to directly apply the slaving principle in synergetics to dig out the order parameters in governing the self-sustained oscillations in this complex system. Recent progresses revealed that some fundamental topologies or sub-networks may dominate the emergence of sustained oscillations. Although the collective self-sustained oscillation emerges from the organization of units in the system, only a small number or some of key units form typical building blocks and play the dominant role in giving rise to collective dynamics. Thus some key topologies that are composed of a minor proportion of units may lead to a collective oscillation and most other units play the role of slaves. We may call these organizing centers the *self-organization core* or the *oscillation source*.

Theoretically, diverse self-sustained oscillatory activities and related determining mechanisms have been reported in different kinds of excitable complex networks. It was discovered that one-dimensional Winfree loops may support self-sustained target group patterns in excitable networks [60, 61]. Moreover, it was also revealed the center nodes and small skeletons to sustain target-wave-like patterns in excitable homogeneous random networks [62–64]. The mechanism of long-period rhythmic synchronous firings in excitable scale-free networks has been explored to explain the temporal information processing in neural systems [65].

On the other hand, node dynamics on biological networks depends crucially on different systems. For example, the gene dynamics is totally different from excitable dynamics, and the network structures are also different. It was found some fundamental building blocks in gene-regulatory networks can support sustained oscillations, and the interesting chaotic dynamics and its mechanism were studied [66–68].

Revealing the key topology of the organizing center, oscillation cores and further the propagation path is the dominant mission in this section. We first propose two useful methods, i.e. the dominant phase-advanced driving method and the functional weight approach, and then apply them to analyze and further pick up the key topologies from dynamics.

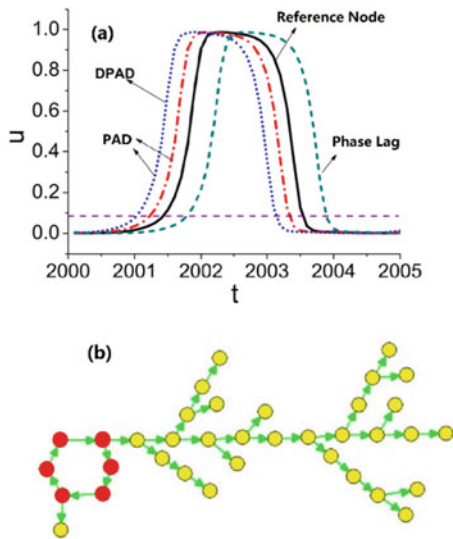
### ***4.3.2 The Dominant Phase-Advanced Driving (DPAD) Method***

An important subject in revealing the coordination of units is to explore the core structure and dynamics in the organization process of a large number of non-oscillatory units. DPAD is a dynamical method that can find the strongest cross driving of the target node when a system is in an oscillatory state. Here, we briefly recall the dynamical DPAD structure [60–64].

Given a network consisting of  $N$  nodes with non-oscillatory local dynamics described by well-defined coupled ordinary differential equations, there are  $M$  ( $M > N$ ) links among these different nodes. We are interested in the situation when the system displays a global self-sustained oscillation and all nodes that are individually non-oscillatory become oscillatory. It is our motivation to find the mechanism supporting the oscillations in terms of the network topology and oscillation time series of each node.

Let us first clarify the significance of nodes in a network with sustained oscillations by comparing their phase dynamics. Obviously, the oscillatory behavior of an individually non-oscillatory node is driven by signals from one or more interactions with advanced phases, if such a phase variable can be properly defined. We call such a signal the *phase-advanced driving* (PAD). Among all phase-advanced interactions, the interaction giving the most significant contribution to the given node can be defined as the *dominant phase-advanced driving* (DPAD). Based on this idea, the corresponding DPAD for each node can be identified. By applying this network reduction approach, the original oscillatory high-dimensional complex network of  $N$  nodes with  $M$  vertices/interactions can be reduced to a one-dimensional unidirectional network of size  $N$  with  $M'$  unidirectional dominant phase-advanced interactions.

An example of clarifying the DPAD is shown in Fig. 4.4a. The black/solid curve denotes the given node as the reference node. Many nodes linking directly to this reference can be checked when a given network is proposed. Suppose there are



**Fig. 4.4** **a** A schematic plot of the DPAD. As comparisons, the reference oscillatory time series, a usual PAD and a phase-lagged node dynamics are also presented, respectively. **b** An example of simplified (unidirectional) network in terms of the DPAD scheme. (Adapted from Ref. [63])

three nodes whose dynamical time series labeled by green/dashed, blue/dotted and red/dash-dotted curves, respectively. The green/dashed curve exhibits a lagged oscillation to the reference curve, therefore one calls it the phase-lagged oscillation. The blue/dotted and red/dash-dotted curves provide the drivings for exciting the reference node, so these curves are identified as PAD. The blue/dotted one presents the earliest oscillation and makes the most significant contribution, thus it is the DPAD.

Figure 4.4b gives an illustration of a DPAD structure consisting of one loop and the nodes outside the loop radiated from the loop. For excitable node dynamics, as shown below, the red nodes form a unidirectional loop that acts as the oscillatory source, and yellow nodes beyond the loop form paths for the propagation of oscillations.

The DPAD structure reveals the dynamical relationship between different nodes. Based on this functional structure, we can identify the loops as the oscillation source, and illustrate the wave propagation along various branches. All the above ideas are generally applicable to diverse fields for self-sustained oscillations of complex networks consisting of individual non-oscillatory nodes.

### 4.3.3 The Functional-Weight (FW) Approach

The above DPAD approach provides a way in analyzing the phase relations embedded in dynamical data to unveil information on unit connections. In some cases, it is possible to get the detailed dynamics of coupled systems. We start from the following dynamical system composed of  $N$  units labeled as  $\mathbf{x} = (x_1, x_2, \dots, x_N)$ :

$$\frac{dx_i}{dt} = \lambda_i x_i + f_i(\mathbf{x}), \quad (4.30)$$

where  $x_i$  denotes the state variable of the  $i$ -th node in a network (for simplicity one adopts the one-dimensional dynamics on the nodes). We separate the linear component from the nonlinear coupling function  $f_i(\mathbf{x})$  at the right-hand side of (4.30). It is our motivation here to explore the topological mechanism of collective oscillation in networks of coupled non-oscillatory units. Therefore the linear coefficients  $\lambda_i$  are negative, and we set  $\lambda_i = -1$  without losing generality.

The functional weight (FW) approach comes from a simple but solidly standing idea: as all the nodes in the networks cannot oscillate individually, the oscillation of any node  $i$  (the target node) is due to the interactions from its neighbors represented by  $f_i(\mathbf{x})$  in Eq. (4.30). However, all the inputs from the neighbors to the target node are mixed together in  $f_i(\mathbf{x})$  by nonlinear functions. To measure the importance of different neighbors to the oscillation of the target node, contributions from all these neighbors should be separated. The cross differential force between the  $i$ -th and the  $j$ -th nodes can be measured by.

$$\delta \dot{f}_{ij}(\mathbf{x}) = (\partial f_i(\mathbf{x}) / \partial x_j) \dot{x}_j, \quad (4.31)$$

where the total driving force is expressed as a simple summation of the contributions of all its neighbors.

$$\delta \dot{f}_i(\mathbf{x}) = \sum_{j=1, j \neq i}^N \delta \dot{f}_{ij}(\mathbf{x}). \quad (4.32)$$

It is emphasized that the differential form  $\partial f_i(\mathbf{x}) / \partial x_j$  rather than  $f_i(\mathbf{x})$  itself plays crucial role in the oscillation generation, because the amount of variation of the target node  $i$  caused by the variation of a given neighbor  $j$  determines the functional driving relationship from  $j$  to  $i$ . At time  $t$ , the weight of the contribution of the  $j$ -th node can be easily computed from Eq. (4.32) as

$$w_{ij}(t) = \frac{|\delta \dot{f}_{ij}(\mathbf{x})|}{\sum_{j=1, j \neq i}^N |\delta \dot{f}_{ij}(\mathbf{x})|} = \frac{|(\partial f_i / \partial x_j) \dot{x}_j|}{\sum_{j=1, j \neq i}^N |(\partial f_i / \partial x_j) \dot{x}_j|}, \quad (4.33)$$

which is nothing but the normalized Jacobian matrix at time  $t$  weighted by  $\dot{x}_j$ . The overall weight is integrated over  $T$  as

$$\bar{w}_{ij} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{t_0}^{t_0+T} w_{ij}(t) dt. \quad (4.34)$$

For periodic oscillatory dynamics,

$$\bar{w}_{ij} = \frac{1}{T} \int_{t_0}^{t_0+T} w_{ij}(t) dt, \quad (4.35)$$

where  $T$  is the period for periodic oscillations. The quantity  $w_{ij}$  introduced here represents the weight of neighbor node  $j$  in driving the target node  $i$  to oscillate and serves as the quantitative measure of the importance of the link from node  $j$  to node  $i$ .  $w_{ij}$  is positive or zero, and normalized as  $\sum_{j=1, j \neq i}^N w_{ij} = 1$ . A zero or small  $w_{ij}$  represents no or a weak functional interaction while large or unity  $w_{ij}$  denotes a strong or dominant driving [66–68].

In the following we will focus on self-sustained oscillations in excitable networks and regulatory gene networks. One can find that both types of systems possess a common feature, that is, only a small number of units participate in the global oscillation, and some fundamental structures act as dominant roles in giving rising to oscillatory behaviors in the system, although the organizing cores differ for these two types of networks.

### 4.3.4 Self-sustained Oscillation in Excitable Networks

#### 4.3.4.1 Oscillation Sources and Wave Propagations

We first take the representative excitable dynamics as a prototype example to reveal the mechanism of global oscillations. Cooperation among units in the system leads to an ordered dynamical topology to maintain the oscillatory process. In regular media, the oscillation core of a spiral wave is a self-organized topological defect. People also found that loop topology is significant in maintaining the self-sustained oscillation. Jahnke and Winfree proposed the dispersion relation in the Oregonator model [68]. Courtemanche et al. studied the stability of the pulse propagation in 1D chains [69]. If a loop composed of excitable nodes can produce self-sustained oscillations, one may call it the *Winfree loop*.

It is not difficult to understand the loop topology for a basic structural basis of collective oscillation for a network of non-oscillatory units. An excitable node in a sustained oscillatory state must be driven by other nodes. To maintain such drivings, a simple choice is the existence of a looped linking among interacting nodes. A local excitation leads to a pulse and are propagated along the loop to drive other nodes in order, which forms a feedback mechanism of repeated driving. Furthermore, the oscillation along the loop can be propagated by nodes outside the loop and spread throughout the system. The propagation of oscillation in the media gives rise to the wave patterns.

The loop structure is ubiquitous in real networks plays an important role in network dynamics. Recurrent excitation has been proposed to be the reason supporting self-sustained oscillations in neural networks. The DPAD method can reveal the underlying dynamic structure of self-sustained waves in networks of excitable nodes and the oscillation source. In complex networks, numerous local regular connections coexist with some long-range links. The former plays an important role in target wave propagation and the latter are crucial for maintaining the self-sustained oscillations.

We use the following Bär-Eiswirth model [60] to describe the excitable dynamics and consider an Erdos–Renyi (ER) random network [29]. The network dynamics is described by.

$$\dot{u}_i = -\frac{1}{\varepsilon}u_i(u_i - 1)\left(u_i - \frac{v_i + b}{a}\right) + D \sum_{j=1}^N A_{i,j}(u_j - u_i), \quad (4.36a)$$

$$\dot{v}_i = f(u_i) - v_i \quad (4.36b)$$

Variables  $u_i(t)$  and  $v_i(t)$  describe the activator and the inhibitor dynamics of the  $i$ -th node, respectively. The function  $f(u)$  takes the following piecewise form:

$$f(u) = \begin{cases} 0, & \text{if } u < 1/3, \\ 1 - 6.75u(u - 1)^2, & \text{if } \frac{1}{3} \leq u < 1, \\ 1, & \text{if } u > 1. \end{cases} \quad (4.37)$$

The relaxation parameter  $\varepsilon \ll 1$  represents the time ratio between the activator  $u$  and the inhibitor  $v$ . The dimensionless parameters  $a$  and  $b$  denote the activator kinetics of the local dynamics and the ratio  $u_T = b/a$  can effectively control the excitation threshold.  $D$  is the coupling strength between linking nodes.  $\mathbf{A} = \{A_{i,j}\}$  is the adjacency matrix. For a symmetric and bidirectional network, the matrix is defined as  $A_{i,j} = A_{j,i} = 1$  if there is a connection linking nodes  $i$  and  $j$ , and  $A_{i,j} = A_{j,i} = 0$  otherwise.

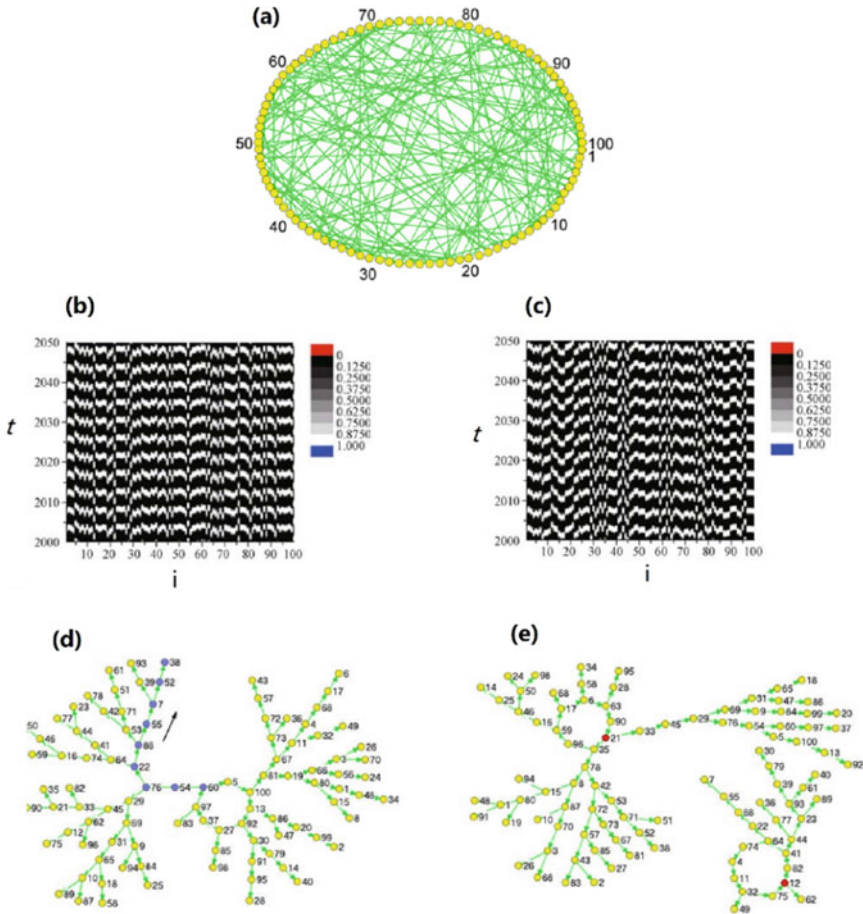
We study the random network shown in Fig. 4.5a as a typical example. Without couplings among nodes, each excitable node is non-oscillatory, i.e. they evolve asymptotically to the rest state  $u = v = 0$  and will stay there perpetually unless some external force drives them away from this state. When a node is kicked from its rest state by a stimulus large enough, the unit can excite by its own internal excitable dynamics.

With the given network structure and parameters, one studies the dynamics of the system by starting from different sets of random initial conditions. The system evolves asymptotically to the homogeneous rest state in many cases. However, one still finds a small portion of tests eventually exhibit global self-sustained oscillations. The spatiotemporal patterns given in Figs. 4.5b, c are two different examples of these oscillatory (both periodic and self-sustained) states.

One can unveil the mechanism supporting the oscillations and the excitation propagation paths by using the DPAD approach. In Figs. 4.5d, e, the reduced directed networks corresponding to the oscillatory dynamics by using the DPAD method are plotted. For the case with dynamics shown in Fig. 4.5b, the single dynamical loop plays the role of oscillation source, with cells in the loop exciting sequentially to maintain the self-sustained oscillation, as shown in Fig. 4.5d. We can observe waves propagating downstream along several tree branches rooted at various cells in the loop. If we plot the spatiotemporal dynamics along these various paths by re-arranging the node indices according to the sequence in the loop, we can find regular and perfect wave propagation patterns. This indicates that the DPAD structure well illustrates the wave propagation paths. For the case with dynamics shown in Fig. 4.5c, the corresponding DPAD structure given in Fig. 4.5e is a superposition of two sub-DPAD paths, i.e. there are two organization loop centers, where two sustained oscillations are produced in two loops and propagate along the trees.

The DPAD structures in Fig. 4.5d, e clearly show the distinctive significance of some units in the oscillation which cannot be observed in Fig. 4.5b, c, where units evolve in the homogeneous and randomly coupled network, and no unit takes any priority over others in topology. Because the unidirectional loop works as the oscillation source, units in the loop should be more important to the contribution of the oscillation.





**Fig. 4.5** **a** An example random network with  $N = 100$  nodes, and each node connects to other nodes with the same degree  $k = 3$ . **(b–c)**: The spatiotemporal evolution patterns of two different oscillatory states in the same network shown in **(a)** by starting from different initial conditions. Both patterns display the evolution of local variable  $u$ . The nodes are spatially arranged according to their indexes  $i$ . **(d)** The DPAD structure corresponding to the oscillation state in **(b)**, where a loop and multiple chains are identified; **e** The DPAD structure corresponding to the oscillation state in **(c)**, where two independent subgraphs are found with each subgraph containing a loop and numerous chains. (Adapted from Ref. [63])

#### 4.3.4.2 Minimum Winfree Loop and Self-sustained Oscillations

Studies on the emergence of self-sustained oscillations in excitable networks indicate that regular self-sustained oscillations can emerge. However, whether there is intrinsic mechanism in determining the oscillations in networks is still unclear. For example, for Erdos–Renyi (ER) networks, whether the connection probability is related to sustained oscillations is an open topic.

In this section we study the occurrence of sustained oscillation depending on the linking probability on excitable ER random networks, and find that the minimum Winfree loop (MWL) is the intrinsic mechanism in determining the emergence of collective oscillations. Furthermore, the emergence of sustained oscillation is optimized at an optimal connection probability (OCP), and the OCP is found to form a one-to-one relationship with the MWL length. This relation is well understood that the connection probability interval and the OCP for supporting the oscillations in random networks are exposed to be determined by the MWL. These three important quantities can be approximately predicted by the network structure analysis, and have been verified in numerical simulations [70].

One adopts the Bär-Eiswirth model (1) on ER networks with  $N$  nodes. Each pair of nodes are connected with a given probability  $P$ , and the total number of connections is  $PN(N - 1)/2$ . By manipulating  $P$ , one can produce a number of random networks with different detailed topologies for a given  $P$ .

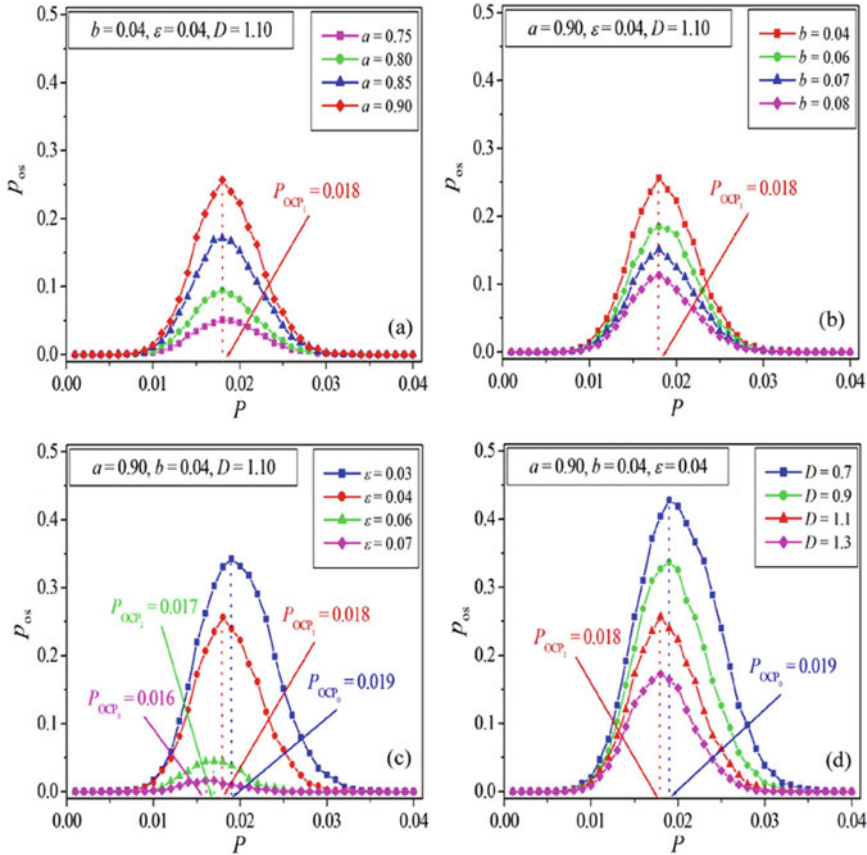
We introduce the oscillation proportion

$$P_{os} = N_{os}/N_{ALL} \quad (4.38)$$

as the order parameter to quantitatively investigate the influence of system parameters on self-sustained oscillations in random networks, where  $N_{ALL}$  is the total number of tests starting from random initial conditions for each set of parameters, and  $N_{os}$  is the number of self-sustained oscillations counted in  $N_{ALL}$  dynamical processes.

In Figs. 4.6a–d, the dependence of the oscillation proportion  $P_{os}$  on the connection probability  $P$  for different parameters  $a$ ,  $b$ ,  $\varepsilon$  and  $D$  on ER random networks with  $N = 100$  nodes are presented. It is shown from all these curves that the system can exhibit self-sustained oscillation in a certain regime of the connection probability, and no oscillations are presented at very small or very large  $P$ . Moreover, an OCP  $P = P_{OCP}$  for supporting self-sustained oscillations can be expected on ER random networks. The number of self-sustained oscillations increases as the parameter  $a$  is increased (see Fig. 4.6a), while  $P_{os}$  decreases as  $b$  is increased as shown in Fig. 4.6b. Moreover, the OCP for supporting self-sustained oscillations is independent of the parameters  $a$  and  $b$ . Figure 4.6c reveals the dependence of  $P_{os}$  on the relaxation parameter  $\varepsilon$ . It is shown from Fig. 4.6c that as  $\varepsilon$  is increased,  $P_{os}$  decreases remarkably. Increasing the coupling strength  $D$  is shown to enhance the sustained oscillation (see Fig. 4.6d).

The non-trivial dependences of collective oscillations on various parameters such as the connection probability  $P$  are very interesting. As discussed above, the excitable wave propagating along an excitable loop can form a 1D Winfree loop, which serves as the oscillation source and maintain self-sustained oscillation in excitable complex networks. Figure 4.7a presents the dependence of the sustained-oscillation period  $T$  of the 1D Winfree loop on the loop length, where a shorter/longer period is expected for a shorter/longer loop length. However, due to the existence of the refractory period of excitable dynamics, a too short 1D Winfree loop cannot support sustained oscillations, implying a minimum Winfree loop (MWL) length  $L_{min}$  for a given set



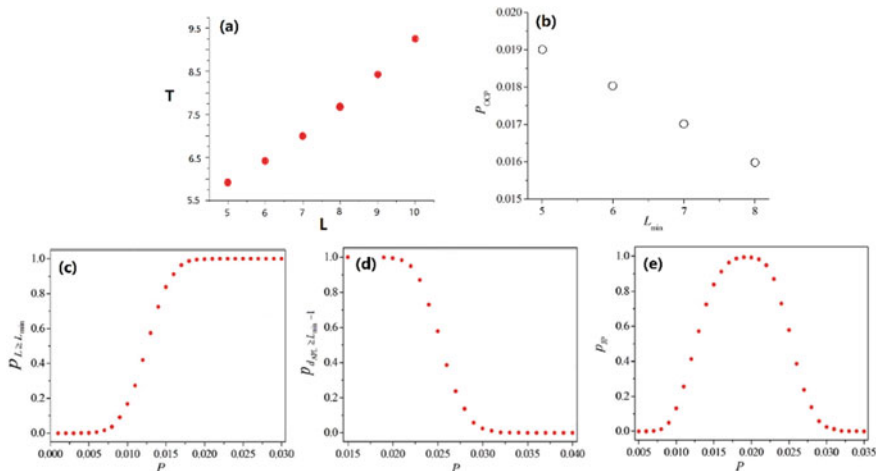
**Fig. 4.6** The dependence of the oscillation proportion  $P_{os}$  on the connection probability  $P$  at different system parameters in excitable ER random networks. The OCPs  $P_{OCP}$  for supporting self-sustained oscillations in ER networks at different parameters are indicated. (Adapted from Ref. [70])

of parameters. Moreover, sustained oscillation ceases for a shortest loop length, corresponding to the MWL.

In Fig. 4.7b,  $P_{OCP}$  is found to build a one-to-one correspondence to  $L_{min}$ , indicating that the emergence of collective oscillations is essentially determined by the MWL. This correspondence can be understood by analyzing the following two tendencies. First, as discussed above, a network must contain a topological loop with a length that is not shorter than the MWL, i.e.

$$L \geq L_{min}.$$

Second, the average path length (APL) of a given network should be large enough so that.



**Fig. 4.7** **a** The dependence of the oscillation period of a Winfree loop against the loop length. When  $L < L_{min}$ , the oscillation ceases. **b** The dependence of the OCP on  $L_{min}$ . **c** The dependence of the proportion of network structures satisfying  $L \geq L_{min}$  on the connection probability  $P$ . **d** The dependence of the proportion of network structures with an APL satisfying  $d_{APL} \geq L_{min} - 1$  on  $P$ . **e** The dependence of the joint probability (JP) on  $P$ . The MWL with the length  $L_{min} = 6$  is used as the example. (Adapted from Ref. [70])

$$d_{APL} \geq L_{min} - 1.$$

These two tendencies propose the necessary conditions for the formation of 1D Winfree loop supporting self-sustained oscillations. Moreover, the first condition leads to the lower critical connection probability, by violating which the network cannot support sustained oscillations. The second condition gives rise to the upper critical connection probability, and if the APL is so short that the loops are too small to support oscillations. In Fig. 4.7c, the probability of the loop length larger than the MWL length is computed against the connection probability  $P$  of an ER network. An increasing dependence can be clearly seen, and a lower threshold exists. As shown in Fig. 4.7d, the probability of an ER network with the APL satisfying  $d_{APL} \geq L_{min} - 1$  is plotted against  $P$ . A decreasing relation and an upper threshold can be found. Necessary condition for sustained oscillation should be a joint probability satisfying both  $L > L_{min}$  and  $d_{APL} \geq L_{min} - 1$ . The dependence of this joint probability on the connection probability  $P$  is the product of Figs. 4.7c, d, which naturally leads to a humped tendency shown in Fig. 4.7e, where an OCP expected for the largest joint probability. This gives a perfect correspondence to the results proposed in Fig. 4.7.

The above discussion indicates that self-sustained oscillations are related to the loop topology and dynamics, and are essentially determined by the MWL. The one-to-one correspondence between the optimal connection probability and the MWL length is revealed. The MWL is the key factor in determining the collective oscillations on ER networks [71–73].

### 4.3.5 Sustained Oscillation in Gene Regulatory Networks

#### 4.3.5.1 Gene Regulatory Networks (GRNs)

Gene regulatory networks (GRNs), as a kind of biochemical regulatory networks in systems biology, can be well described by coupled differential equations (ODEs) and have been extensively explored in recent years. The ODEs describing biochemical regulation processes are strongly nonlinear and often have many degrees of freedom. We are concerned with the common features and network structures of GRNs.

Different from the above excitable networks, positive feedback loops (PFLs) and negative feedback loops (NFLs) have been identified in various biochemical regulatory networks and found to be important control modes in GRNs [46, 50–53]. Self-sustained oscillation, bi-rhythmicity, bursting oscillation and even chaotic oscillations are expected for these objects. Moreover, oscillations may occur in GRNs with a small number of units. The study of sustained oscillations in small GRNs is of great importance in understanding the mechanism of gene regulation processes in very large-scale GRNs. Network motifs as subgraphs can appear in some biological networks, and they are suggested to be elementary building blocks that carry out some key functions in the network. It is our motivation to unveil the relation between network structures and the existence of oscillatory behaviors in GRNs.

We consider the following GRN model:

$$\frac{dp_i}{dt} = f_i(\mathbf{p}) - p_i, \quad (4.39)$$

where  $\mathbf{p} = (p_1, p_2, \dots, p_N)$ , and the function  $f_i(\mathbf{p})$  satisfies the following form:

$$f_i(\mathbf{p}) = \begin{cases} A_i(\mathbf{p}) & \text{for active regulation only} \\ R_i(\mathbf{p}) & \text{for repressive regulation only} \\ A_i(\mathbf{p})R_i(\mathbf{p}) & \text{for joint regulation} \end{cases} \quad (4.40)$$

The active regulation function is.

$$A_i(\mathbf{p}) = act_i^h / (act_i^h + K^h), \quad (4.41a)$$

and the repressive regulation function is written as

$$R_i(\mathbf{p}) = K^h / (rep_i^h + K^h), \quad (4.41b)$$

where

$$act_i = \sum_{(j=1)}^N \alpha_{ij} p_j, \quad rep_i = \sum_{(j=1)}^N \beta_{ij} p_j, \quad i, j = 1, 2, \dots, N,$$

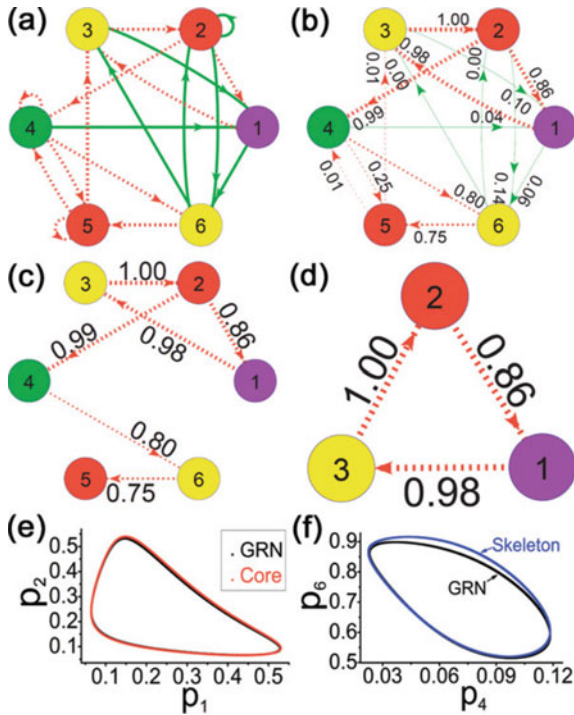
and  $p_i$  is the expression level of gene  $i$ ,  $0 < p_i < 1$ . The adjacency matrices  $\alpha$ ,  $\beta$  determine the network structure, which are defined in such a way that  $\alpha_{ij} = 1$  if gene  $j$  activates gene  $i$ ,  $\beta_{ij} = 1$  if gene  $j$  inhibits gene  $i$  and  $\alpha_{ij} \cdot \beta_{ij} = 0$  for no dual-regulation of gene  $i$  by gene  $j$ .  $act_i(rep_i)$  represents the sum of active (repressive) transcriptional factors to node  $i$ . The regulated expression of genes is represented by Hill functions with cooperative exponent  $h$  and the activation coefficient  $K$ , characteristic for many real genetic systems [66].

### 4.3.5.2 Skeletons and Cores

Based on the known oscillation data of GRNs, we can explicitly compute all the functional weights  $w_{ij}$  between any two genes based on Eqs. (4.33)–(4.35) and draw the FW maps for the oscillatory networks. Figure 4.9a shows a simple example of 6-node GRN with the dynamics given by Eq. (4.39), and computer simulations indicate that this GRN possesses an oscillatory attractor for a set of parameters given in Ref. [66]. From the oscillation data, we can draw the FW map in Fig. 4.9b. An interesting as well as valuable feature of Fig. 4.9b is that the  $w_{ij}$  distribution is strongly heterogeneous, *i.e.* some weights  $w_{ij} \approx 1$  indicate a significant control while many others  $w_{ij} \approx 0$  represent weak functional links. The heterogeneity allows us to explore the self-organized functional structures supporting oscillatory dynamics by the strongly weighted links. In Fig. 4.9c, we remove all the less important links with  $w_{ij} \leq w_{th}$ , where the threshold value  $w_{th} = 0.30$ . By removing any interaction and node in the network, we always mean to delete their oscillatory parts while keep the average influence of the deleted parts as control parameters in Eq. (4.34). The reduced subnetwork shown in Fig. 4.9c retaining only strongly weighted interactions is called the *skeleton* of the oscillatory GRN. Furthermore, we can reduce the skeleton network by removing the nodes without output one by one, and finally obtain an irreducible subnetwork where each node has both input and output which is defined as the *core* of the oscillation.

For the network Fig. 4.8b the core is given in Fig. 4.8d. In Fig. 4.8e, f, we plot the oscillation orbits of the original GRN and those of the core and skeleton in 2D ( $p_1, p_2$ ) and ( $p_4, p_6$ ) phase planes, respectively. It is found that the dynamical orbits of the original network given in Fig. 4.8a can be well reproduced by the reduced structures in Fig. 4.8c, d, conforming that the skeleton and core defined by the highly weighted links can dominate the essential dynamics of the original GRN. Considering Fig. 4.8d–e, c–f it is clear that the core Fig. 4.8d serves as the oscillation source of the GRN while the skeleton Fig. 4.8c plays the role of main signal propagation paths from the core throughout the network.

The analysis presented in Fig. 4.8 can be well applied to more complicated GRNs with larger numbers of nodes and links to fulfill the reduction of network complexity for understanding and controlling network dynamics. Numerical works for a large network with many nodes and links indicates that the comparisons of the dynamics of the core and skeleton with those of the original network are in good agreement.



**Fig. 4.8** An example of the functional-weight (FW) map for an oscillatory GRN. **a** An oscillatory GRN with  $N = 6$  genes and  $I = 15$  links. Different node colors represent different phases of the oscillatory nodes. Green full and red dotted arrowed lines represent active and repressive interactions, respectively. **b** The FW map of all interactions computed by Eqs. (3.34) and (3.35). **c** The reduced interaction skeleton obtained from (b) by deleting links with small weights. **d** An irreducible core structure obtained by only retaining interactions with both input and output in (c), which serves as the oscillation source propagating through the path of the skeleton (c). **e** A comparison of a dynamical orbit of the original GRN (a) with that of core (d) in a 2D phase plane. **f** A comparison of an orbit of the GRN (a) with that of skeleton (c) in another 2D phase plane. Good agreements between the orbits of the full GRN and that of the reduced subnetworks can be found

Let us give a brief summary of the above discussions on the emergence of sustained oscillation dynamics in networks of non-oscillatory units. In this case, the feedback topology is the fundamental mechanism. For a very large network, not every node or link contributes to the collective oscillation, and only a small portion of nodes and their connections dominates, forming some fundamental building blocks such as motifs, loops, or cores. To dig out these basic topologies, an appropriate topology reduction is the key point. This meanwhile leads naturally to dynamics reduction, revealing the emergence of self-organized oscillation from collaborations of non-oscillatory node dynamics. It is a significant issue to make a topology reduction to reduce the dimensionality of dynamics of a complex system to obtain the essential structural ingredient of emergence behaviors. In recent years we developed the related techniques and ways in revealing the embedded topologies [74, 75].

It should be stressed that there exist many fundamental topologies on a large network that may induce various possible sustained oscillations. Competitions and collaborations among these sub-networks and rhythms lead to complicated dynamics.

## 4.4 Synchronization: Cooperations of Rhythms

### 4.4.1 *Synchrony: An Overview*

Collective behavior of complex and nonlinear systems has a variety of specific manifestations, and synchronization should be one of the most fundamental phenomena [76, 77]. Discussions of fundamental problems in synchronization cover a variety of fields of natural science and engineering, and even many behaviors in social science are closely related to the basic feature of synchronization. Many specific systems such as pendula, musical instruments, electronic devices, lasers, biological systems, neurons, cardiology and so on, exhibit very rich synchronous phenomena.

Let us first focus on fireflies, a magical insect on earth. A male firefly can release a special luminescent substance to produce switched periodic flashes to attract females. This corresponds to a typical biological sustained oscillation. A surprising phenomenon occurs when a large number of male fireflies gather together in the dark of night to produce synchronized flashings. This phenomenon was first written in the log of the naturalist and traveler, Engelbert Kaempfer in 1680 when he traveled in Thailand. Later in 1935, Hugh Smith reported his observation on the synchronized flashing of fireflies [78]. More explorations have been performed after these initial observations. Interestingly, Buck elaborated on this phenomenon and published two articles with the same title “synchronized flashing of fireflies” on the same journal in 1938 and 1988, respectively [79].

It is a complex issue in biology to explore the implication of the synchronization flashing of male fireflies. The physical mechanism behind this phenomenon may be more important and interesting because it is a typical dynamical process. If each firefly is considered as an oscillator, the synchronous flashing is an emergent order produced by these oscillators. This collective behavior is obviously originated from the interaction between individuals, and the interplay between them results in an adjustment of the rhythm of every firefly.

Accidentally and interestingly, before Kempfer’s finding of firefly synchronization, Huygens described the synchronization of two coupled pendula [80] in 1673. He discovered that a couple of pendulum clocks hanging from a common support had synchronized, i.e. their oscillations coincided perfectly and the pendula moved always in opposite directions. He also proposed his understanding of the phenomenon by attributing to the coupling brought by the beam that hang two pendula and the coupling-induced energy transfer between the two pendulum clocks.



Although synchronization behaviors had been found in different disciplines such as physics, acoustics, biology, and electronic devices, a common understanding embedded these seemingly distinct phenomena was still lack. A breakthrough was the study of sustained oscillations and limit cycles in nonlinear systems in the early twentieth century [44]. Further, it is theoretically important to study the synchronization between the limit cycles of driving or interaction on the basis of limit cycles [76, 77].

A fruitful modelling of synchronization was pioneered by Winfree, who studied the nonlinear dynamics of a large population of weakly coupled limit-cycle oscillators with distributed intrinsic frequencies [81]. The oscillators can be characterized by their phases, and each oscillator is coupled to the collective rhythm generated by the whole population. Therefore, one may use the following equations of motion to describe the dynamical evolution of interacting oscillators:

$$\dot{\theta}_i = \omega_i + \left( \sum_{j=1}^N X(\theta_j) \right) Z(\theta_i), \quad (4.42)$$

where  $j = 1, \dots, N$ . Here  $\theta_i$  denotes the phase of the  $i$ -th oscillator,  $\omega_i$  its natural frequency. Each oscillator  $j$  exerts a phase-dependent influence  $X(\theta_j)$  on all the other oscillators. The corresponding response of oscillator  $i$  depends on its phase through the sensitivity function  $Z(\theta_i)$ .

Winfree discovered that such a population of non-identical oscillators can exhibit a remarkable cooperative phenomenon in terms of the mean-field scheme. When the spread of natural frequencies is large compared to the coupling, the system behaves incoherently, with each oscillator running at its natural frequency. As the spread is decreased, the incoherence persists until a certain threshold is crossed, i.e. then a small cluster of oscillators freezes into synchrony spontaneously.

Kuramoto put forward Winfree's intuition about the phase model by adopting the following universal form [82]:

$$\dot{\theta}_i = \omega_i + \sum_{j=1}^N \Gamma_{ij}(\theta_j - \theta_i), \quad (4.43)$$

where the coupling functions  $\Gamma$  depends on the phase difference and can be calculated as integrals involving certain terms from the original limit-cycle model. A tractable phase model of (4.43) was further proposed by adopting a mean-field sinusoidal coupling function:

$$\dot{\theta}_i = \omega_i + \frac{K}{N} \sum_{j=1}^N \sin(\theta_j - \theta_i), \quad (4.44)$$

$K \geq 0$  is the coupling strength. The frequencies are randomly chosen from a given probability density  $g(\omega)$ , which is usually assumed to be one-humped and symmetric about its mean  $\omega_0$ . This mean-field model was heretofore called the Kuramoto model.

The Kuramoto mean-field model can be successfully solved by using the self-consistency approach in terms of statistical physics, which reveals that a large number of coupled oscillators can overcome the disorder due to different natural frequencies by interacting with each other, and the synchronized state emerges in the system.

The success of Winfree's and Kuramoto's works aroused extensive studies of synchronization under more generalized cases (interested reader may refer the review papers and the monographs [83–85]). The study of coupled phase oscillator synchronization, and the Kuramoto model on complex networks has become the focus of research [86–89].

Apart from the self-consistency approach, recently Ott and Antonsen proposed an approach (OA ansatz) to obtain the dynamical equations of order parameters [90, 91]. Strogatz explained the physical meaning of the OA ansatz based on the Watanabe-Strogatz transformation [92, 93].

In recent years, with the widespread studies of chaotic oscillations, the notion of synchronization has been generalized to chaotic systems [76]. The study of synchronization of coupled chaotic oscillators extended the scope of synchronization dynamics, and different types of chaos synchronization such as complete/identical synchronization, generalized synchronization, phase synchronization, and measure synchronization were revealed [88, 94, 95].

## 4.4.2 *Microdynamics of Synchronization*

Let us begin with the simplest scenario to explore the synchronous dynamics. It is very important to discuss the microscopic mechanism of synchronization, which can make us better understand how a large number of coupled oscillators form ordered behaviors through interaction and self-organization.

### 4.4.2.1 **Phase-Locking of Two Limit-Cycle Oscillators**

We consider two mutually coupled oscillators  $z_{1,2}(t)$  that are described by Eq. (4.28) but with different natural frequencies. They are coupled to each other and obey the following dynamical equations of motion

$$\dot{z}_{1,2} = (\lambda_{1,2} + i\omega_{1,2})z_{1,2} - b_{1,2}z_{1,2}|z_{1,2}|^2 + K_{1,2}z_{2,1}|z_{2,1}|^2, \quad (4.45)$$

where  $\lambda_{1,2} > 0$ ,  $K_{1,2} > 0$ .  $b_{1,2}$  are two real parameters, and  $\omega_1 \neq \omega_2$  are the natural frequencies of the two oscillators. The third term at the right hand side of

Eq. (4.45) represents the interaction between two oscillators. By introducing the polar coordinates  $r_{1,2}(t)$  and  $\theta_{1,2}(t)$  as

$$z_{1,2}(t) = r_{1,2}(t)e^{i\theta_{1,2}(t)}, \quad (4.46)$$

the motion in two complex equations of (4.45) can be decomposed into four real equations of the amplitudes and phases as

$$\dot{r}_{1,2}(t) = \lambda_{1,2}r_{1,2}(t) - b_{1,2}r_{1,2}^3 + K_{1,2}r_{2,1}^3 \operatorname{Re}[e^{i(\theta_{2,1}-\theta_{1,2})}], \quad (4.47a)$$

$$\dot{\theta}_{1,2}(t) = \omega_{1,2} + K_{1,2}(r_{2,1}^3/r_{1,2}) \operatorname{Im}[e^{i(\theta_{2,1}-\theta_{1,2})}]. \quad (4.47b)$$

It can be seen from (4.47b) that the coupling term will adapt the actual phase velocities  $\dot{\theta}_{1,2}(t)$  even if two oscillators have different natural frequency  $\omega_{1,2}$ .

We consider the possibility of the attracting tendency of two oscillators in the presence of coupling. By comparing the coupling terms in (4.47a) and (4.47b), it can be found that the first one is of order  $r^3$ , while the latter is of order  $r^2$ . Therefore the coupling term in (4.47a) can be neglected, and the two equations in (4.47a) are decoupled and can be solved. As  $t \rightarrow \infty$ ,  $r_{1,2} \rightarrow r_{10,20} = \sqrt{\lambda_{1,2}/b_{1,2}}$ . By substituting the amplitudes  $r_{1,2}$  in (4.47b), one can obtain the following coupled phase equations:

$$\dot{\theta}_{1,2}(t) = \omega_{1,2} + K_{1,2} \left( \frac{r_{20,10}^3}{r_{10,20}} \right) \sin(\theta_{2,1} - \theta_{1,2}). \quad (4.48)$$

The above procedure implies the separation of time scales of the amplitude and phase, which is actually the consequence of the slaving principle by setting  $\dot{r}_{1,2} = 0$ , i.e. the amplitudes  $r_{1,2}(t)$  are fast state variables, while the phases  $\theta_{1,2}(t)$  are neutrally-stable slow variables. Let us keep in working out the phase dynamics of two coupled oscillators. By introducing the phase difference  $\theta(t) = \theta_2(t) - \theta_1(t)$  and natural-frequency difference  $\Omega = \omega_2 - \omega_1$ , Eq. (4.48) can be changed to

$$\Delta\dot{\theta} = \Omega - \alpha \sin \Delta\theta, \quad (4.49)$$

where the parameter

$$\alpha = K_2 \left( \frac{r_{10}^3}{r_{20}} \right) + K_1 \left( \frac{r_{20}^3}{r_{10}} \right).$$

The evolution of (4.49) can be easily solved by integrating the differential equations and eventually one gets

$$t = \int_{\bar{\varphi}_0}^{\bar{\varphi}} d\Delta\theta / [\Omega - \alpha \sin \Delta\theta]. \quad (4.50)$$

when  $|\Omega| > |\alpha|$ , the integral (4.50) can be worked out in one period of the phase, and the corresponding period is expressed as

$$T = \int_0^{2\pi} \frac{d\Delta\theta}{\Omega - \alpha \sin \Delta\theta}, \quad (4.51)$$

i.e. the phase difference evolves periodically with the period  $T$ :

$$\Delta\dot{\theta} = \Delta\dot{\theta}_2 - \Delta\dot{\theta}_1 = \omega'_2 - \omega'_1 \approx 2\pi/T, \quad (4.52)$$

i.e. the actual frequency difference of two coupled oscillators depends on the integral (5.51). When  $\alpha \ll \Omega$ , one has

$$\omega'_2 - \omega'_1 \approx \Omega = \omega_2 - \omega_1. \quad (4.53)$$

On the other hand, when  $|\Omega| < |\alpha|$ , the integral (5.51) diverges at  $\sin \Delta\theta_0 = \alpha/\Omega$ . This implies that as  $t \rightarrow \infty$ ,  $\Delta\theta$  tends to a fixed value  $\Delta\theta_0 = \arcsin(\alpha/\Omega)$ , and the period  $T \rightarrow \infty$  in (5.51). From (5.52) one has

$$\Delta\dot{\theta} = \dot{\theta}_2 - \dot{\theta}_1 \rightarrow 0, \quad (5.54)$$

i.e. the frequencies of the two oscillators are pulling to each other and eventually locked. In fact, because  $\alpha$  is proportional to the coupling strength, the critical condition  $|\Omega| \leq |\alpha|$  means that coupling strength  $K_{1,2}$  should be strong enough to overcome the natural-frequency difference  $\Omega$ . Therefore, the critical condition is  $\alpha_c = \omega_2 - \omega_1$ . Near this critical point, i.e. one has  $\langle \Delta\dot{\theta} \rangle \sim (\alpha_c - \alpha)^{1/2}$ , where  $\langle \cdot \rangle$  represents a long-time average. This implies a saddle-node bifurcation at the onset of synchronization of two coupled oscillators.

As an inspiration, one finds from the above study of synchronization between two interacting limit-cycle oscillations that: (1) The phase is the dominant degree of freedom in the process of synchronization of coupled oscillators as compared to the amplitude variable; (2) The coupling function between oscillators is typically of the sinusoidal form of the phase difference. These two points are in agreement with the proposition of Winfree and Kuramoto in modelling synchronization, which is also a very important starting point in describing the synchronization problem.

#### 4.4.2.2 Synchronization Bifurcation Tree

For  $N > 2$  coupled oscillators, it is not an easy task to give analytical discussions of the microdynamics of synchronization. One usually performs numerical simulations and compute some useful quantities. Let us consider the following nearest-neighbor coupled oscillators:

$$\dot{\theta}_i = \omega_i + \frac{K}{3}[\sin(\theta_{i+1} - \theta_i) + \sin(\theta_{i-1} - \theta_i)], \quad (4.55)$$

where  $i = 1, 2, \dots, N$ ,  $\{\omega_i\}$  are natural frequencies of oscillators,  $K$  is the coupling strength. Without losing generality, we assume that  $\sum_i \omega_i = 0$ .

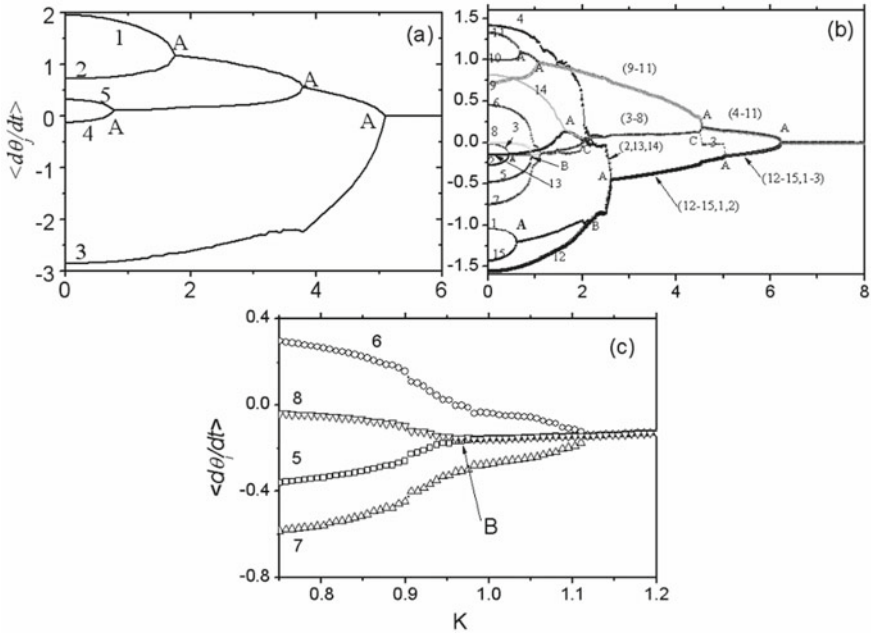
When the coupling strength  $K$  is increased from 0, different from the two-oscillator case, the system will show complicated synchronization dynamics because of the competition between the ordering induced by the coupling and the disorder of natural frequencies. For the nearest-neighbor coupling case, there is an additional competition, i.e. the competition between the coupling distance and the natural-frequency differences. As the coupling strength increases, the system will gradually reach the *global synchronization*. There is a critical coupling  $K_c$ , when  $K > K_c$  the frequencies of all the oscillators are locked to each other. As  $K < K_c$ , a portion of oscillators are synchronized, which is called *partial synchronization*. To observe the synchrony process, we define the average frequency of the  $i$ -th oscillator as

$$\bar{\omega}_i = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \dot{\theta}_i(t) dt. \quad (4.56)$$

Synchronization between the  $i$ -th and the  $j$ -th oscillators is achieved when  $\bar{\omega}_i = \bar{\omega}_j$ . As the coupling strength changes, oscillators will undergo a coordinated process to achieve global synchronization.

To observe the synchronization of multiple oscillators clearly, we introduced the so-called *synchronization bifurcation tree (SBT)*, which is defined as the set of the relation  $\{\bar{\omega}_i(K)\}$ , i.e. the relationship of the average frequencies of all oscillators and the coupling strength  $K$ . The SBT method gives a tree-structured process of synchronization transitions and exhibits vividly how oscillators are organized to become synchronized by varying the coupling [96–99].

In Fig. 4.9a–b, we plot the average frequencies  $\{\bar{\omega}_i\}$  defined in Eq. (4.56) against the coupling strength  $K$  for  $N = 5$  and 15, respectively, by varying  $K$  from  $K = 0$  to  $K = K_c$ . In both figures, we find interesting transition trees of synchronizations. When  $K = 0$ , all oscillators have different winding numbers, and an increase of the coupling may lead to a merging of  $\bar{\omega}_i$ . As two oscillators become synchronous with each other, their frequencies become the same at a critical coupling and keep the same (a single curve) with further increase of the coupling strength.

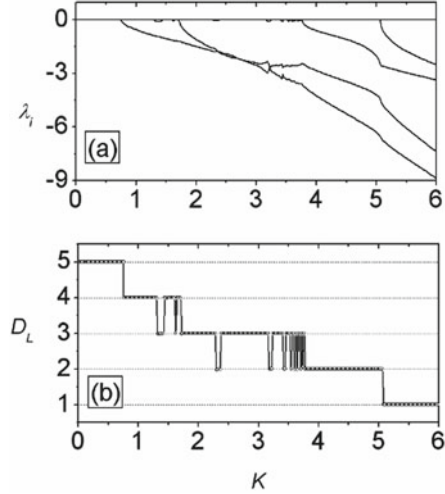


**Fig. 4.9** Transition trees of synchronization for averaged frequencies of oscillators versus the coupling  $K$ . **a**  $N = 5$ ; **b**  $N = 15$ . Note the existence of three kinds of transitions labeled A, B, and C. **c** An enlarged plot of the nonlocal phase synchronization for  $N = 15$ . (Adapted from Ref. [98])

An interesting behavior of SBT is the clustering of oscillators, i.e. several synchronous clusters can be formed with the increase of the coupling, and these clusters have different frequencies and numbers of oscillators. Clusters also form into larger clusters by reducing the number of clusters. For sufficiently strong coupling, only few clusters (usually two clusters) are kept and eventually merge into a single synchronous cluster. The formation of a single cluster implies the global synchronization of all oscillators. For both the SBT in Fig. 4.9a, b, one can observe the interesting tree cascade of synchrony.

There are many ways in investigating the dynamics of a system, among which the most convincing tool is the computation of Lyapunov exponents. If one gets the Lyapunov-exponent spectrum (LES)  $\{\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N\}$  of the system, the basic properties of the attractor can be well traced and understood. By observing the variation of the LES with system parameters, it is instructive to understand the relation between changes of the synchronous dynamics and attractor transitions with parameters. When there is one or more Lyapunov exponents larger than zero, the motion of the system is chaotic. If there are  $M \geq 2$  zero exponents and no positive exponents, then the motion of system is quasi-periodic, i.e. the attractor in phase space is an  $M$ -dimensional torus (labeled as  $T^M$ ). The Ruelle-Takens quasiperiodic route to chaos and the structural instability of the high-dimensional torus is a very important topic. In fact, in some cases high-dimensional torus can also survive with

**Fig. 4.10** **a** The variation of the LES with the coupling strength for  $N = 5$  coupled oscillators with the same parameters as in Fig. 4.9a. **b** The corresponding Lyapunov dimension  $D_L$  varying with the coupling strength  $K$ . Steplike behavior can be observed. (Adapted from Ref. [98])



a non-zero measure. We can find the existence of high-dimensional quasiperiodicity in the weak-coupling regime.

In Fig. 4.10a, we calculated the variation of the LES with the coupling strength for  $N = 5$  coupled oscillators with the same parameters as in Fig. 4.9a. It can be seen from Fig. 4.10a that when  $K \leq 0.75$ , all the exponents  $\lambda_{1\sim 5} = 0$ , indicating that the high-dimensional quasiperiodic motion is on  $T^5$ , and this 5-dimensional torus keeps stable with  $K$  changes in a large scale. When  $K > 0.75$ , one can see that one zero Lyapunov exponent becomes negative, and the number of zero exponents is reduced by one, indicating that the dynamical attractor has a topological transition from the torus  $T^5$  to  $T^4$ . This transition implies a bifurcation. By comparing the SBT shown in Fig. 4.9a, we can see that this transition point corresponds to the synchrony between oscillator  $i = 4$  and  $i = 5$ . If we keep increasing the coupling strength  $K$ , we can further find the more transitions with each critical point corresponding to a zero Lyapunov exponent becoming negative, and also corresponding to the synchrony of oscillators. Therefore, we assert that the synchronization process of coupled periodic oscillators is accompanied with the process of dynamical transitions from high-dimensional quasiperiodic to low-dimensional quasiperiodic motions. In the vicinity of each critical point, the negative Lyapunov exponent  $\lambda_i$  satisfies the following scaling law:

$$\lambda_i \propto -A(K_c^i - K)^{1/2}, \quad (4.57)$$

where  $A$  is a factor, and  $K_c^i$  is the synchronous critical point.

The transition to a lower-dimensional torus means the reduction of the dimension of the attractor in phase space during the synchronization process [97, 98]. We can compute the dimensions of the attractor varying with the coupling strength  $K$ . A

simple method for calculating the attractor dimension is the *Kaplan-Yorke conjecture* that the dimension can be obtained from the LES:

$$D_L = M + \frac{1}{|\lambda_{M+1}|} \sum_{j=1}^M \lambda_j, \quad (4.58)$$

here  $M$  is an integer that satisfies the following criteria:

$$\sum_{j=1}^M \lambda_j \geq 0, \quad \sum_{j=1}^{M+1} \lambda_j < 0. \quad (4.59)$$

The Lyapunov exponents here are arranged in the descent order, i.e.  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ . We call the quantity  $D_L$  the *Lyapunov dimension*. In Fig. 4.10b, we calculated the Lyapunov dimension  $D_L$  varying with the coupling strength  $K$  for  $N = 5$ , where the steplike behavior can be clearly observed. The system keeps an integer dimension  $D_L$  until a synchronization transition occurs, and then the dimension jumps by one to another integer value. We can also see the upside-down jumps between two integers, which suggests that many quasiperiodic windows embedded in high-dimensional tori are still observable.

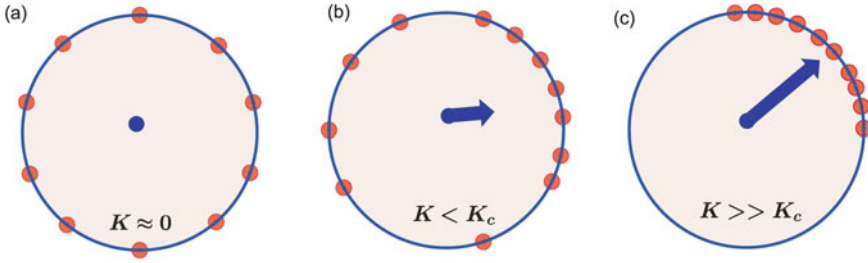
Here we showed the importance of microscopic synchronous dynamics of coupled oscillators. With the increase of coupling strength, a large number of coupled oscillators undergo a cascade of transitions from partial to global synchronizations, and this process is accompanied with the decrease of the phase-space dimension. When all oscillators reach the global synchronization, the dynamics of the system falls into a very low-dimensional manifold in phase space. This means that at the onset of synchronization, only a few variables are required to characterize the synchronization dynamics of coupled oscillators. This fact provides a foundation for the macroscopic description of synchronization of coupled oscillators.

### 4.4.3 Kuramoto Model: Self-Consistency Approach

It is not practical to give all the details of synchronization among oscillators when the population is very large. To measure the coherent behavior of oscillators, it is more convenient to introduce the following *order parameter*, also called the *coherence factor*, which is defined as the mean-field average of the complex functions of phases as

$$\alpha_1 = Re^{i\Theta} = \frac{1}{N} \sum_{j=1}^N e^{i\theta_j}. \quad (4.60)$$





**Fig. 4.11** A schematic process of the synchronization transition from incoherent to coherent states, where oscillators are labeled as dots on the unit circle, and the arrow corresponds to the order parameter  $\alpha_1$ . **a** For a weak coupling  $K \approx 0$ , oscillators are not synchronized, and their phases are evenly distributed in  $[0, 2\pi]$ , the order parameter  $\alpha_1 \approx 0$ ; **b** With the increase of coupling strength, more and more oscillators will be synchronized and no longer evenly distributed,  $|\alpha_1| \neq 0$ ; **c** As the coupling becomes very large, oscillators form a single synchronized cluster, and  $|\alpha_1|$  becomes larger (longer length of the arrow)

Here  $R$  is the modulus of the complex order parameter  $\alpha_1$ , which describes the degree of coherence of oscillators, and  $\Theta$  is a collective phase.

Although the natural frequencies of oscillators are different, interactions among them can organize to an ordered state. By varying the coupling strength from 0, the actual frequencies of oscillators  $\{\bar{\omega}_i\}$  defined in (4.56) will shift from their natural values  $\{\omega_i\}$  and move closer to each other, which has been well exhibited in microdynamics shown by SBT. When the coupling strength is very weak, only oscillators with natural frequencies very close to each other can be synchronized, but their proportions can be almost ignored as  $N \gg 1$ , almost all oscillators are evenly spaced within  $0 \sim 2\pi$  at any time, as schematically shown in Fig. 4.11a. In this case it can be easily verified that  $R = 0$  when asynchronous state prevails. With the increase of coupling strength, more and more oscillator will be synchronized and their average frequencies  $\Omega_i$  become equal. The phases of synchronized oscillator will be locked and keep close, i.e. the phases of locked oscillators are no longer evenly distributed, as shown in Fig. 4.11b. When all  $\{\Omega_i\}$  are equal, the order parameter  $R$  will become non-zero, indicating that oscillators will maintain a fixed phase relationship at the onset of synchrony. It has been proved that there exists a critical coupling strength  $K_c$ , when  $K \leq K_c$  the order parameter  $R = 0$ , while  $R \neq 0$  as  $K \geq K_c$ . For stronger couplings, phases of oscillators become closer to form a compact synchronized group, as shown in Fig. 4.11c. In thermodynamic limit  $N \rightarrow \infty$ , the above transition from asynchronous to synchronous states is a typical nonequilibrium phase transition at the critical point  $K_c$ . The Kuramoto model is a featured system that can be analytically solved and exhibit the above phase transition.

In the study of this synchronization transition, an important task is to determine the critical coupling strength of synchronous transitions  $K_c$  and the order parameters  $R$ . The mean-field coupling of the Kuramoto model gives us the chance to deal with in terms of the self-consistency method, if it is possible to build the equation of order parameter as a function of coupling strength.

The self-consistency approach is based on the assumption that the system has a stationary state that does not change over time. In this stationary state, the order parameter is a time-independent quantity, which can be obtained by its definition and the equations of the motion. The self-consistency method is not limited to specific dynamics and is one of the widely used methods in coupled oscillator systems.

When the number of oscillators  $N \gg 1$ , the order parameter defined by (4.60) is irrelevant to  $N$  and does not change with time. Due to the mean-field feature of the summation in the coupling term, by using (4.60) one can rewrite the Kuramoto model (4.44) to the following form:

$$d\theta_i/dt = \omega_i + KR \sin(\Theta - \theta_i). \quad (4.61)$$

If one regards  $R$  as a parameter, Eq. (4.61) indicates that dynamics of all oscillators are decoupled, i.e. the influence of other oscillators on the  $i$ -th oscillator is described by the parameter  $R$ . If one knows  $R$ , (4.61) can be well worked out. However,  $R$  is also an undetermined coefficient. A possible solution is to build an equation of  $R$ . This equation is just the self-consistency equation [82, 84, 85]. Thus how to get the self-consistency equation becomes the core task.

When  $N \gg 1$ , we do not care about the micro-states  $\{\theta_i(t)\}$  any more, but how are these phases distributed at any time  $t$ . Let  $\rho(\theta, \omega, t)$  denote a dynamical variable representing the number density of the oscillators with natural frequency  $\omega$  and phase  $\theta$  at time  $t$ . Since we are interested in the thermodynamic limit  $N \rightarrow \infty$ , we expect that an infinitely large number of oscillators will fall into an arbitrarily small but finite interval  $\Delta\theta$ . The single-oscillator distribution function  $\rho(\theta, \omega, t)$  depends not only on the phase  $\theta$  variable, but also on the natural frequency  $\omega$ .  $\rho(\theta, \omega, t)$  is  $2\pi$ -periodic and satisfies the normalization condition

$$\int_{-\infty}^{+\infty} \int_0^{2\pi} \rho(\theta, t) d\theta d\omega = 1. \quad (4.62)$$

The distribution of these oscillator phases directly determines the relative average. The order parameter introduced in (4.60) as the summation of all oscillators can be replaced by

$$\alpha_1 = R e^{i\Theta} = \int_{-\infty}^{+\infty} \int_0^{2\pi} e^{i\theta} \rho(\theta, t) d\theta d\omega. \quad (4.63)$$

We are mainly interested in the behavior, especially the long-term behavior of  $\rho(\theta, \omega, t)$ . Since dynamical Eq. (4.44) is invariant under a translation of  $\theta_i \rightarrow \theta_i + \theta_0$ , one expects that the simplest collective behavior may be described by a uniform and stationary distribution:  $\rho(\theta, \omega, t) = 1/2\pi$ , i.e.  $\{\theta_i\}$  are uniformly distributed in the range  $0 \sim 2\pi$ . It can be easily verified that the case of  $R = 0$  corresponds to the incoherent state. This is always a solution of the system, but it is not always

stable. When  $K \geq K_c$  the uniformly distributed solution is unstable and replaced by the collective solution where the oscillators are locked, i.e. the solution when  $\Omega_i$  is equal, at which point the oscillator can maintain a fixed phase, and  $R \neq 0$ . In the latter case, all oscillators are oscillatory with the same frequency  $\bar{\omega}$ ,  $\Theta = \bar{\omega} t$ . By introducing variables

$$\phi_i = \theta_i - \bar{\omega}t, \quad (4.64)$$

Equation (4.61) can be written as

$$d\phi_i/dt = \omega_i - \bar{\omega} - KR \sin \phi_i. \quad (4.65)$$

Equation (4.65) has the following two types of solutions:

- (1) **Synchronous Solution:** When Eq. (4.65) that describes satisfies

$$|\omega_i - \bar{\omega}| \leq KR, \quad (4.66)$$

the solution for the phase  $\phi_i$  is the fixed point

$$\phi_i = \sin^{-1}[(\omega_i - \bar{\omega})/KR]. \quad (4.67)$$

This means that the  $i$ -th oscillator oscillates with the frequency  $\bar{\omega}$ , and all oscillators that satisfy the conditions (4.66) will oscillate with the same frequency  $\bar{\omega}$ , i.e. oscillators satisfying (4.66) of course will be in a synchronous state.

- (2) **Asynchronous solution:** when Eq. (4.65) satisfies

$$|\omega_i - \bar{\omega}| > KR \quad (4.68)$$

the phase  $\phi_i$  is an oscillatory solution. Because it represents the phase difference, oscillators with natural frequencies satisfying (4.68) are not synchronized.

The above discussion indicates that coupled oscillators can naturally divide into the synchronous and asynchronous groups according the conditions (4.66) and (4.68). In the following we will focus their contributions to the distribution  $\rho(\theta)$  and the order parameter  $R = |\alpha_1|$ .

If the Kuramoto system is driven by an external noise, the model is dynamically described by

$$\dot{\theta}_i = \omega_i + \frac{K}{N} \sum_{j=1}^N \sin(\theta_j - \theta_i) + \xi_i(t), \quad (4.69)$$

where the noise is usually assumed to be a spatiotemporally uncorrelated Gaussian white noise that satisfies

$$\langle \xi_i(t) \rangle = 0, \quad \langle \xi_i(t) \xi_j(t') \rangle = D \delta_{ij} \delta(t - t'), \quad (4.70)$$

where  $D$  is the noise intensity. By introducing the order parameter  $R$ , Eqs. (4.69) can be written as

$$d\theta_i/dt = \omega_i + KR \sin(\Theta - \theta_i) + \xi_i(t), \quad (4.71)$$

The distribution function  $\rho(\theta, \omega, t)$  satisfies the Fokker–Planck equation

$$\frac{\partial \rho}{\partial t} = -\frac{\partial(v\rho)}{\partial \theta} + D \frac{\partial^2 \rho}{\partial \theta^2}, \quad (4.72)$$

where

$$v = \omega + KR \sin(\Theta - \theta). \quad (4.73)$$

In the absence of noise ( $D = 0$ ), Eq. (4.72) is reduced to the continuity equation satisfying the phase distribution function:

$$\frac{\partial \rho}{\partial t} = -\frac{\partial(v\rho)}{\partial \theta}. \quad (4.74)$$

Because  $\phi = \theta - \bar{\omega} t$ , the distribution function  $\rho(\phi, \omega, t)$  satisfying equations can be easily derived from the equation of  $\rho(\theta, \omega, t)$ .

Considering the natural-frequency distribution of oscillators, if one is only concerned with the distributions of the phase  $\theta$  or  $\phi$ , the reduced distribution can be obtained by averaging over the natural frequencies as

$$\rho(\phi, t) = \int \rho(\omega, \phi, t) g(\omega) d\omega, \quad (4.75)$$

In the following discussions we mainly discuss the synchronous transition in the absence of noises and the case of stationary phase distributions.

The above two types of solutions enlighten us that the stationary distribution  $\rho(\phi)$  can be decomposed into the synchronous and asynchronous parts:

$$\rho(\phi) = \rho_s(\phi) + \rho_{as}(\phi). \quad (4.76)$$

The synchronous part includes the oscillators with phases  $\phi_i$  being fixed points, thus  $\rho_s(\phi)$  can be obtained by natural frequencies that satisfies  $d\phi_i/dt = 0$ , i.e.  $\omega = \bar{\omega} + KR \sin \phi$ . Thus one has

$$\rho_s(\phi) = g(\omega) \left| \frac{d\omega}{d\phi} \right| = KR g(\bar{\omega} + KR \sin \phi) \cos \phi, \quad \phi \in \left[ -\frac{\pi}{2}, \frac{\pi}{2} \right]. \quad (4.77)$$

For those oscillators that are not synchronized, the phases  $\{\phi_i\}$  change with time. Because  $\phi_i$  evolves non-uniformly over time, the probability that the phase being within  $\phi \rightarrow \phi + d\phi$  at time  $t$  should be inversely proportional to the phase velocity  $|\dot{\phi}|$ , i.e.

$$\rho(\phi, \omega) \propto |\dot{\phi}|^{-1}. \quad (4.78)$$

This can also be obtained from the steady state of Eq. (4.74) by setting  $\partial\rho(\phi, \omega, t)/\partial t = 0$ , leading to  $\partial(v\rho)/\partial\phi = 0$ , and thus  $\rho \propto v^{-1} = |\dot{\phi}|^{-1}$ . By substituting the equation of motion (4.65) and making the normalization, one gets

$$\begin{aligned} \rho(\phi, \omega) &= \left\{ |\omega - \bar{\omega} - KR \sin \phi| \int_0^{2\pi} \frac{d\phi}{|\omega - \bar{\omega} - KR \sin \phi|} \right\}^{-1} \\ &= \frac{\sqrt{(\omega - \bar{\omega})^2 - (KR)^2}}{2\pi |\omega - \bar{\omega} - KR \sin \phi|}. \end{aligned} \quad (4.79)$$

Because oscillators with  $|\omega - \bar{\omega}| > KR$  are not synchronized, by summing up all oscillators satisfying this frequency condition, one obtains

$$\rho_{as}(\phi) = \int_{-\infty}^{\bar{\omega}-KR} g(\omega)\rho(\phi, \omega)d\omega + \int_{\bar{\omega}+KR}^{\infty} g(\omega)\rho(\phi, \omega)d\omega \quad (4.80)$$

By introducing  $x = \omega - \bar{\omega}$  and considering the symmetry property of the function  $g(\omega)$

$$g(\bar{\omega} + x) = g(\bar{\omega} - x), \quad (4.81)$$

The asynchronous part of the distribution can be written as

$$\rho_{as}(\phi) = \int_{KR}^{\infty} \frac{g(\bar{\omega} + x)x\sqrt{x^2 - (KR)^2}}{\pi[x^2 - (KR \sin \phi)^2]} dx. \quad (4.82)$$

The order parameter can be rewritten as

$$Re^{i\Theta} = \int_{-\pi}^{\pi} e^{i(\phi + \bar{\omega}t)} \rho(\phi) d\phi = \int_{-\pi}^{\pi} e^{i\phi + i\bar{\omega}t} [\rho_s(\phi) + \rho_{as}(\phi)] d\phi. \quad (4.83)$$

Because the asynchronous part  $\rho_{as}(\phi)$  has the even term  $\sin^2\phi$ , the integral of  $\rho_{as}(\phi)$  in (4.83) is zero. Hence only the symmetric part  $\rho_s(\phi)$  contributes to the

integral (4.83). By further separating the real and imaginary parts of the integral (note that the  $R$  is real) one obtains

$$R = KR \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^2 \phi g(\bar{\omega} + KR \sin \phi) d\phi, \quad (4.84)$$

$$0 = KR \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos \phi \sin \phi g(\bar{\omega} + KR \sin \phi) d\phi. \quad (4.85)$$

From (4.85) one can obtain the average frequency  $\bar{\omega}$ . (4.84) is a typical self-consistency equation of  $R$ , which can be used to determine both  $R$  and the critical coupling strength  $K_c$ . When  $K \geq K_c$ ,  $R$  changes from 0 to a very small value. Near the critical point,  $R \ll 1$ ,  $g(\omega)$  can be expanded into Taylor series as

$$g(\bar{\omega} + KR \sin \phi) \approx g(\bar{\omega}) + \frac{g''(\bar{\omega})}{2} (KR)^2 \sin^2 \phi + O(R^4), \quad (4.86)$$

$$g''(\bar{\omega}) = \left. \frac{d^2 g(\omega)}{d\omega^2} \right|_{\omega=\bar{\omega}}. \quad (4.87)$$

Substituting the expansion to (4.84) one obtains

$$1 = \frac{\pi K}{2} g(\bar{\omega}) - \frac{1}{16} \pi K^3 R^3 g''(\bar{\omega}) + O(R^4). \quad (4.88)$$

when  $K \rightarrow K_c$ ,  $R \rightarrow 0$ , the second and third terms in (4.88) approach zero, one may determine the critical coupling strength as [84]

$$K_c = 2/[\pi g(\bar{\omega})]. \quad (4.89)$$

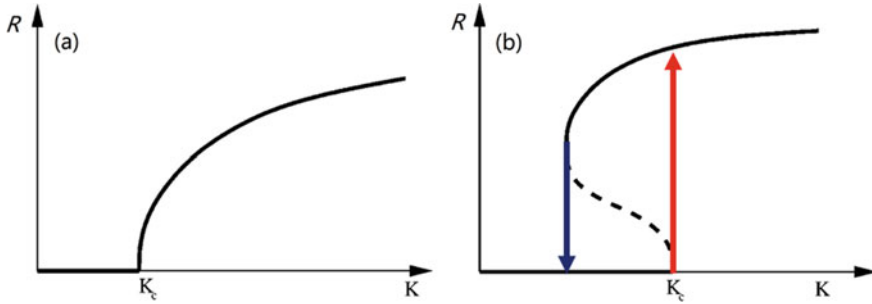
Putting (4.89) back to (4.88), the critical behavior near  $K_c$  can be determined as

$$R \approx \sqrt{\frac{8g(\bar{\omega})(K - K_c)}{g''(\bar{\omega})K^3}}. \quad (4.90)$$

It can be seen from (4.89) and (4.90) that the characteristics of the natural-frequency distribution  $g(\omega)$  around  $\bar{\omega}$  is very important. For example, for the Lorentz distribution with the form

$$g(\omega) = \{\pi[(\omega - \bar{\omega})^2 + \gamma^2]\}^{-1} \gamma, \quad (4.91)$$

the results above are simplified to



**Fig. 4.12** The order parameter varying with the coupling strength  $K$ . **a**  $g''(\bar{\omega}) > 0$ . **b**  $g''(\bar{\omega}) < 0$ . (Adapted from Ref. [85])

$$K_c = 2\gamma, \quad (4.92a)$$

$$R = \sqrt{(1 - 2\gamma/K)}. \quad (4.92b)$$

One can see that the order parameter  $R$  exhibits a typical continuous phase transition behavior

$$R \propto (K - K_c)^{1/2} \quad (4.93)$$

near the critical point, as shown in Fig. 4.12a. This implies a typical second-order phase transition similar to statistical physics. It should be pointed out that this is a kind of non-equilibrium phase transition, which is the result of the system ordering (coupling) overwhelming the disorder brought by natural frequency random distribution.

Usually the second derivative in Eq. (4.90) at  $\omega = \bar{\omega}$  should satisfy  $g''(\bar{\omega}) > 0$ , which means that the distribution function  $g(\omega)$  is unimodal. In this case one expects the second-order phase transition as shown in Fig. 4.12(a). If  $g''(\bar{\omega}) < 0$ , i.e. the distribution function is not a unimodal one, then near the critical point one has

$$R \propto (K_c - K)^{1/2}, \quad (4.94)$$

as shown by the dashed line in Fig. 4.12b, indicating that the synchronized state is unstable and practically unobservable. There exists a solid line in Fig. 4.12b representing the stable synchronous branch, and the transitions between the asynchronous and synchronous states are discontinuous due to the existence of unstable branch between two stable branches in the bifurcation diagram. Therefore the transition to synchronization when  $g''(\bar{\omega}) < 0$  is the first-order phase transition [85], and the emergence of the bistable regime indicates the hysteresis behavior as one varies the coupling strength upwardly and downwardly. We will no longer discuss the details

of this case. In fact, the stable synchronous branch in the bifurcation diagram can no longer be obtained by the self-consistency method.

#### 4.4.4 Order Parameter Dynamics: Equations of Motion

For a large number of interacting oscillators, we can deal with its dynamics at the statistical or macroscopic level instead of tedious microscopic details. The self-consistency approach successfully predicts the transition to synchronization at the macroscopic level, but it requires that the order parameter is time independent. Further studies revealed in many cases the order parameter is far from stationary. Therefore it is necessary to study the non-stationary dynamics of the order parameter.

Another interesting issue is the emergent process of order parameters. According to the principle of synergetics, the emergence of an order parameter is the result of the spontaneous collaboration and competition among various degrees of freedom with different time scales. Why does  $\alpha_1$  becomes dominant and acts as the order parameter to characterize the macroscopic behavior of the system instead of other quantities? A general coupled oscillator system is usually complicated, and it is important to set up a theoretical framework of order parameter dynamics. In this section, we focus on the emergence of dominant order parameters in terms of the basic idea of slaving principle [6, 21, 100, 101].

Let us consider  $N$  fully coupled oscillators, and the equations of motion can be written as

$$\dot{\theta}_j(t) = F(\boldsymbol{\alpha}, \theta_j, \boldsymbol{\beta}, \gamma_j), \quad (4.95)$$

$j = 1, 2, \dots, N$ . Here  $\boldsymbol{\beta} = \{\beta_1, \beta_2, \dots\}$  represent a set of uniform control parameters, i.e. these parameters are the same for all oscillators.  $\boldsymbol{\gamma} = \{\gamma_1, \gamma_2, \dots, \gamma_N\}$  is a set of non-uniform control parameters, these parameters are not the same for different oscillators. A typical example is that the natural frequencies of oscillators are usually different, and in this case  $\{\gamma_i = \omega_i, i = 1, 2, \dots, N\}$ .

We define the following set of collective parameters  $\boldsymbol{\alpha} = \{\alpha_n\}$ :

$$\alpha_n = \frac{1}{N} \sum_{j=1}^N e^{in\theta_j}. \quad (4.96)$$

Obviously  $\alpha_1$  has the same expression as that defined in (4.60). Parameters  $\alpha_n$  with  $n > 1$  are high-order parameters. We call this set of parameters  $\boldsymbol{\alpha}$  the generalized order parameters.

In thermodynamic limit  $N \rightarrow \infty$ , the detailed dynamical information of an individual oscillator is no longer relevant. It is more convenient to introduce density distribution function  $\rho(\boldsymbol{\gamma}, \theta, t)$  and study the statistical property of the coupled oscillator system, where  $\rho(\boldsymbol{\gamma}, \theta, t)d\theta$  is the probability that the phase of an oscillator falls



in  $\theta \rightarrow \theta + d\theta$  at time  $t$ . The evolution of the distribution function corresponding to Eq. (4.95) is typically the continuity equation similar to (4.74)

$$\partial\rho/\partial t + \partial(\rho v)/\partial\theta = 0, \quad (4.97)$$

where the phase velocity is  $v = F(\boldsymbol{\alpha}, \theta, \boldsymbol{\beta}, \boldsymbol{\gamma})$ .

By considering the inhomogeneity in the system, the total distribution function should sum over the non-identical parameter  $\boldsymbol{\gamma}$ ,

$$\rho(\theta, t) = \sum_i \rho(\gamma_i, \theta, t). \quad (4.98)$$

For example, when we study coupled oscillators with non-uniform natural frequencies, the total distribution function can be derived by summing over the natural frequency as

$$\rho(\theta, t) = \int \rho(\omega, \theta, t)g(\omega)d\omega. \quad (4.99)$$

where  $g(\omega)$  is the distribution function of natural frequencies.

The above discussion establishes the statistical description of coupled oscillators. The phase distribution function contains all the information of the collective behavior of coupled oscillators. As long as one solves the equation of the distribution function (97), all the other statistical and macroscopic quantities can be calculated in terms of the distribution function. As for the synchronization of coupled oscillators, we are concerned with the order parameters, which can well describe the degree of synchronization. By using the distribution function, the expression of the order parameters in terms of the sum of  $e^{in\theta}$  can be expressed as the following integral for a homogeneous system:

$$\alpha_n = \int e^{in\theta} \rho(\theta, t)d\theta. \quad (4.100)$$

In the presence of heterogeneity, one also needs to sum over the non-uniform parameters. For example, if the non-uniform parameter is the natural frequency, then the integral form of the generalized order parameter can be expressed as

$$\alpha_n = \int e^{in\theta} \rho(\omega, \theta, t)g(\omega)d\omega d\theta. \quad (4.101)$$

The above expression indicates explicitly that the order parameters  $\alpha_n$  are actually the statistical average, or  $n$ -th order moment of the phase factor  $e^{in\theta}$ . Note that statistically the description of all orders of moments is equivalent to that of the distribution function, i.e. one can obtain the complete information from the other party.

In addition, we can see from the above two expressions that the generalized order parameters are actually the Fourier coefficients of the distribution function  $\rho(\theta, t)$ .

The order parameters  $\alpha_n$  can describe different orderness of a population of oscillators. First, it can be easily seen that  $|\alpha_n| \leq 1$ .  $|\alpha_n| = 0$  represents a statistically homogeneous and random distribution of oscillators with phases in  $[0, 2\pi]$ , which is denoted as the incoherent state. When  $|\alpha_n| = 1$ , oscillators are in the clustering state. For example,  $|\alpha_1| = 1$  corresponds to the state that all oscillators possess the same phase  $\{\theta_i(t) = \theta(t), i = 1, 2, \dots, N\}$ , i.e. the globally synchronous state.  $|\alpha_2| = 1$  refers to a two-cluster state, where oscillators are divided into two synchronous clusters with a phase shift  $\pi$ :  $\{\theta_i(t) = \theta(t), \theta_j(t) = \theta(t) + \pi, i = 1, 2, \dots, M, j = M + 1, \dots, N\}$ . Therefore  $|\alpha_n| = 1$  represents the state of  $n$  synchronous clusters with phase shift  $2\pi/n$ , and the state with order parameter  $0 < |\alpha_n| < 1$  is the partially synchronous state.

We first consider the case of  $N$  coupled identical phase oscillators. Our mission is to derive the evolution dynamics of the order parameters. For a finite number of oscillators, the dynamics of the generalized order parameters  $\alpha$  should be equivalent to that of the microdynamics oscillators, whose equations of motion are written as

$$\dot{\theta}_j(t) = F(\alpha, \theta_j, \beta), j = 1, 2, \dots, N. \quad (4.102)$$

By taking the time derivative of the order parameter (4.96) on both sides and inserting the equations of motion (4.102), one gets

$$\dot{\alpha}_n = \frac{in}{N} \sum_{j=1}^N e^{in\theta_j} F(\alpha, \theta_j, \beta). \quad (4.103)$$

Because  $F(\alpha, \theta, \beta)$  is a  $2\pi$ -periodic functions of the phase  $\theta$ , it can be expanded into the Fourier series as

$$F(\alpha, \theta_j, \beta) = \sum_{k=-\infty}^{\infty} f_k(\alpha, \beta) e^{ik\theta_j}. \quad (4.104)$$

The real function  $F(\alpha, \theta, \beta)$  requires the Fourier coefficients satisfy  $f_{-k}(\alpha, \beta) = \overline{f_k(\alpha, \beta)}$ , where  $\overline{f_k}$  is the complex conjugate of  $f_k$ . Inserting the expansion (4.104) into (4.103) and by using the definition of  $\alpha_n$ , one may get the equations of motion for the order parameters:

$$\dot{\alpha}_n = in \sum_{k=-\infty}^{\infty} f_k(\alpha, \beta) \alpha_{k+n}. \quad (4.105)$$

It can be found that dynamical behavior of  $\alpha_n(t)$  depends on all the other order parameters  $\{\alpha_{k+n}\}$ . Let us consider a truncated case of the system (4.102), i.e. only the first-order order expansion of the coupling function (4.104) is kept, then

$$F(\boldsymbol{\alpha}, \theta) = f_1(\boldsymbol{\alpha})e^{i\theta} + f_{-1}(\boldsymbol{\alpha})e^{-i\theta} + f_0(\boldsymbol{\alpha}), \quad (4.106)$$

and the order parameter Eq. (4.105) can be greatly simplified. By substituting (4.106) into the order parameter motion Eq. (4.105) one obtains

$$\dot{\alpha}_n = in \left[ f_1(\boldsymbol{\alpha})\alpha_{n+1} + \bar{f}_1(\boldsymbol{\alpha})\alpha_{n-1} + f_0(\boldsymbol{\alpha})\alpha_n \right], \quad (4.107)$$

where  $n \geq 0$ ,  $\alpha_{-n} = \bar{\alpha}_n$ .

One should point out that the solution of (4.105) possesses the same difficulty as that of the original dynamical Eqs. (4.102). In fact, these two sets of equations are equivalent descriptions, and the generalized order parameters can be regarded as a set of collective variables transformed from the phase variables. This set of coupled order parameter equations are difficult to solve. On the other hand, one can seek for some specific solutions. One trivial solution is the incoherent/asynchronous solution  $\alpha_n \equiv 0$ .

When the coupling strength is increased, oscillators will synchronize to each other, and the microscopic motion will collapse into a low-dimensional phase space. In the generalized order-parameter space, the system will also collapse to a low-dimensional space. Therefore, according to the slaving principle, only a few (or even only one) of these generalized order parameters will survive to be the dominant order parameters, and other parameters will be fast variables and can be adiabatically eliminated [20, 21]. Here we will explore this interesting question from a theoretical perspective.

## 4.4.5 Emergence of Order Parameters

### 4.4.5.1 The Ott-Antonsen (OA) Ansatz

The above simple three-diagonal iterative form of the order-parameter dynamical equations implies the symmetry of the system and possible low-dimensional dynamics. One of the simplest possibilities is the case when there exists a certain relationship between different order parameters. A possible scenario is that all higher-order parameters  $\{\alpha_{n \geq 2}\}$  depend on  $\alpha_1$ . This can be easily understood since our previous self-consistency approach to the Kuramoto model is based on the study of  $\alpha_1$ . By assuming a uniform form of function representing this dependence, let us look for the following trial solution:

$$\alpha_n = G(\alpha_1, n), \quad (4.108)$$

where  $G(\alpha_1, n)$  is a differentiable function. The time derivative of (4.108) leads to

$$\dot{\alpha}_n = G_{\alpha_1}(\alpha_1, n)\dot{\alpha}_1. \quad (4.109)$$

where  $G_{\alpha_1}(\alpha_1, n) = \partial G(\alpha_1, n)/\partial \alpha_1$ . By using (4.107), one obtains

$$in(f_1\alpha_{n+1} + \bar{f}_1\alpha_{n-1} + f_0\alpha_n) = iG_{\alpha_1}(f_1\alpha_2 + \bar{f}_1\alpha_0 + f_0\alpha_1). \quad (4.110)$$

The coefficients of the Fourier expansion terms  $f_i$  or  $\bar{f}_i$  at both sides in (4.110) are equal, leading to the following relations:

$$n\alpha_{n+1} = G_{\alpha_1}\alpha_2, \quad n\alpha_{n-1} = G_{\alpha_1}\alpha_0, \quad n\alpha_n = G_{\alpha_1}\alpha_1.$$

Because  $\alpha_0 = \int \rho(\omega, \theta, t)g(\omega)d\omega d\theta = 1$ , one obtains

$$G_{\alpha_1} = \frac{n\alpha_{n+1}}{\alpha_2} = n\alpha_{n-1} = \frac{n\alpha_n}{\alpha_1}. \quad (4.111)$$

This naturally leads to

$$nG = G_{\alpha_1}\alpha_1, \quad (4.112)$$

One thus obtains the following form of the function:

$$\alpha_n = G(\alpha_1, n) = \alpha_1^n. \quad (4.113)$$

By inserting (4.113) to (4.107), one finally gets the equation of motion of  $\alpha_1$  as

$$\dot{\alpha}_1 = i[f_1(\alpha_1)\alpha_1^2 + \bar{f}_1(\alpha_1) + f_0(\alpha_1)\alpha_1]. \quad (4.114)$$

It is interesting that (4.113) is just the ansatz recently proposed by Ott and Antonsen [90, 91], which was called the **Ott-Antonsen (OA) ansatz** thereafter. Therefore, an infinite-dimensional dynamical system is reduced to a two-dimensional order parameter equation. Undoubtedly this is a great reduction and simplification of a complex system.

#### 4.4.5.2 Poisson Invariant Manifold

In fact, Ott and Antonsen proposed the above ansatz based on the distribution function. The distribution function can be expanded into Fourier series, where the expansion coefficients are the generalized order parameters  $\alpha_n$ , that is,

$$\rho(\theta, t) = \frac{1}{2\pi} \left[ 1 + \sum_{n=1}^{\infty} (\bar{\alpha}_n(t)e^{in\theta} + \alpha_n(t)e^{-in\theta}) \right]. \quad (4.115)$$

Generally, the above summation can be executed to get the distribution function only when all the Fourier coefficients  $\{\alpha_n\}$  are known. Ott and Antonsen assumed that the coefficients  $\{\alpha_n\}$  are not independent of each other, and they are all determined

by the complex function  $\alpha_1(t)$  and satisfy the following power law, namely

$$\alpha_n(t) = \alpha_1^n(t), \quad \bar{\alpha}_n(t) = \bar{\alpha}_1^n(t). \quad (4.116)$$

One then gets the following form:

$$\rho(\theta, t) = \frac{1}{2\pi} \left[ 1 + \sum_{n=1}^{\infty} (\bar{\alpha}_1^n(t) e^{in\theta} + \alpha_1^n(t) e^{-in\theta}) \right], \quad (4.117)$$

By satisfying the above power-law relation, the summation (4.117) on the right side is simply the power series and can be worked out to obtain the Poisson summation form of the distribution:

$$\rho(\theta, t) = \frac{1}{2\pi} \frac{1 - r^2}{1 - 2r \cos(\theta - \Theta) + r^2}, \quad (4.118)$$

where  $r$  and  $\Theta$  are the amplitude and the phase of the parameter  $\alpha_1(t) = r e^{i\Theta}$ , respectively.

The result (4.118) indicates that the phase distribution is completely determined by  $\alpha_1(t)$ . Because the solution satisfying the relation (4.113) obeys a degenerate equation of motion, the relation (4.113) will be satisfied in the evolution process. The order parameter  $\alpha_1(t)$  may vary with time, and this consequently leads to the change of the distribution  $\rho(\theta, t)$  with the evolution of the system. However, it can be easily found from that (4.118) that  $\rho(\theta, t)$  always keeps an invariant form of the Poisson summation. If the initial phase density distribution of the system  $\rho(\theta, t = 0)$  satisfies the Poisson-summation distribution, then the density distribution  $\rho(\theta, t)$  will always keep this property. Therefore the order-parameter relation (4.113) and the degenerate equation of motion (4.114) are also called the **invariant Poisson-summation sub-manifold** of the dynamical system (4.107). An important feature of this invariant manifold is that  $\alpha_1(t)$  can be either time dependent or time independent.

#### 4.4.5.3 The Inhomogeneous Case

Let us further consider the case of coupled non-identical oscillators. We consider the natural frequency of individual oscillator as the inhomogeneous parameter. Assuming that the natural frequencies obey the distribution  $G(\omega)$ . The dynamical equations of the system can be written as

$$\dot{\theta}_j(t) = F(\boldsymbol{\alpha}, \theta_j, \boldsymbol{\beta}, \omega_j), \quad j = 1, 2, \dots, N \quad (4.119)$$

when  $N \rightarrow \infty$ , the density function  $\rho(\omega, \theta, t)$  can be introduced, and the generalized order parameter is written as

$$\alpha_n(t) = \int \alpha_n(\omega, t)g(\omega)d\omega, \quad (4.120)$$

where  $\alpha_n(\omega, t)$  is the local order parameter with the natural frequency  $\omega$ :

$$\alpha_n(\omega, t) = \int e^{in\theta} \rho(\omega, \theta, t)d\theta, \quad (4.121)$$

and it can also be understood as the  $n$ -order Fourier expansion of  $\rho(\omega, \theta, t)$ . The following recursive equation can be obtained using the continuity equation

$$\dot{\alpha}_n(\omega, t) = in \sum_{j=-\infty}^{\infty} f_j(\alpha, \beta, \omega)\alpha_{j+n}(\omega, t), \quad (4.122)$$

Similar to (4.106), if the coupling function contains only the first-order Fourier coefficients, then the Ott-Antonsen ansatz for a given value of the natural frequency  $\omega$

$$\alpha_n(\omega, t) = \alpha_1^n(\omega, t), \quad (4.123)$$

is a set of special solutions of Eqs. (4.111). Specifically, the first order parameter is

$$\alpha_1(t) = \int \alpha_1(\omega, t)g(\omega)d\omega. \quad (4.124)$$

When the distribution function  $g(\omega)$  is the rational fraction of  $\omega$ , such as Lorentz distribution, one can analytically extend the real  $\omega$  to the complex regime. If there is no divergence ( $|\alpha_1(t)| \leq 1$ ), the evolution equation of  $\alpha_1(t)$  can be worked out.

#### 4.4.5.4 The Mean-Field Kuramoto Model Revisited

Now let us study the synchronization transition of the mean-field coupled oscillator systems in terms of the Ott-Antonsen ansatz by adopting the classical Kuramoto model (4.44) as an example. By using (4.124) and letting  $z(t) = \alpha_1(t)$ , one can rewrite the Fourier components in the coupling function (4.104) as

$$f_1 \equiv -Kz/2i, \quad \bar{f}_1 \equiv K\bar{z}/2i, \quad f_0 \equiv \omega. \quad (4.125)$$

This means that the coupling function of the Kuramoto model only contains the first-order Fourier components shown by (4.106). Then by using (4.114) one can get

$$\dot{\alpha}_1(\omega, t) = \frac{1}{2}[2i\omega\alpha_1(\omega, t) + Kz(t) - K\bar{z}(t)\alpha_1^2(\omega, t)]. \quad (4.126)$$

If the natural-frequency distribution function  $g(\omega)$  is the following Lorentz form

$$g(\omega) = 1/[\pi(\omega^2 + 1)], \quad (4.127)$$

then by inserting it into (4.124) one has

$$z(t) = \int_{-\infty}^{\infty} \frac{\alpha_1(\omega, t)d\omega}{\pi(\omega^2 + 1)}. \quad (4.128)$$

This integral can be done by extending to the complex plane of  $\omega$ . Obviously the extension should be to the upper half complex plane. By using the Cauchy residue theorem, one obtains

$$z(t) = \alpha_1(\omega = i, t). \quad (4.129)$$

By substituting (4.129) to (4.126)–(4.128), one has

$$\dot{z}(t) = \frac{1}{2}z(t)(K - 2 - K|z(t)|^2). \quad (4.130)$$

This is a complex dynamical equation, which describes a two-dimensional real-space dynamics. It can be seen that there exists a critical point  $K_c = 2$  for the dynamical system (4.130). When  $K \leq K_c$ , the system has the only one fixed-point solution  $z \equiv 0$ , which denotes the asynchronously disordered state. When  $K \geq K_c$ , the zero solution becomes unstable, and the system experiences a bifurcation to the non-zero branch

$$|z| = \sqrt{(K - K_c)/K}. \quad (4.131)$$

This non-zero solution represents the emergence of the self-organized synchronization state. By comparing with the result (4.93) obtained in terms of the self-consistency approach proposed in Sec. 4.3, one can find that here we obtain the same results.

An important issue is the stability of the OA invariant manifolds [92, 93]. Recently, we studied the stability of the OA manifold by using the analysis in the functional space of the phase distribution function [102, 103]. We proved that the OA manifold is in fact the two-dimensional invariant manifold in the infinite-dimensional density functional space. This greatly expands the applicability of the OA ansatz as an effective method for analyzing the dynamics of globally coupled phase oscillators [100], where the OA approach has its validity.

Studies on synchronization of the globally coupled phase oscillators can be naturally extended to the case of complex networks:

$$\dot{\theta}_i = \omega_i + \sum_{j=1}^N K_{ij} \sin(\theta_j - \theta_i), \quad i = 1, 2, \dots, N \quad (4.132)$$

This can be considered by considering the coupling weights among oscillators that depend on their natural frequencies:

$$K_{ij} = \begin{cases} K|\omega_i|/N, & \text{for IN - coupling case} \\ K|\omega_j|/N, & \text{for OUT - coupling case} \end{cases} \quad (4.133)$$

When all weights are identical, i.e.  $K_{ij} = K/N$ , the discussion returns to the usual Kuramoto model. Both the self-consistency approach and the OA ansatz can be applied to discussions of this type of globally coupled oscillator systems [104–107]. One can obtain a generalized formula of the critical coupling for synchronization. However, the order parameter dynamics exhibits complicated bifurcations and non-stationary dynamics.

#### 4.4.6 Synchronizations on Star Networks

As an application of the above dimension-reduction scheme of phase dynamics of coupled oscillators, let us study the synchronization of coupled phase oscillators on star networks [108–111]. The star network topology has the typical heterogeneity property, which is very important in studies of collective dynamics on scale-free networks. In a heterogeneous network, such as a scale-free network, hub plays a dominant role. Hence a star motif with a central hub is a typical topology in grasping the essential property of the heterogeneous networks. It has been revealed that an abrupt transition, namely the explosive synchronization, can take place on scale-free networks, which means a large number of oscillators that evolve incoherently can suddenly become synchronous into a large-size cluster at a critical coupling strength [112–114]. The key point in understanding this discontinuous synchronization transition is the dynamical analysis of the multi-stability of miscellaneous synchronous attractors in phase space [115, 116]. However, it is difficult to get an analytical insight in a high-dimensional phase space. Recently, we have revealed the mechanism of synchronization transition by analyzing the collective dynamics in a low-dimensional complex order parameter space in terms of the above dimension-reduction approach [48, 108].

##### 4.4.6.1 The Star-Networked Phase Model

By adopting oscillators on  $N$  leaf nodes with frequencies  $\{\omega_j\}$  and the hub with  $\omega_h$ , the equations of motion can be written as



$$\begin{aligned}\dot{\theta}_h &= \omega_h + K \sum_{j=1}^N \sin(\theta_j - \theta_h - \alpha), \\ \dot{\theta}_j &= \omega_j + K \sin(\theta_h - \theta_j - \alpha), j \in [1, N],\end{aligned}\quad (4.134)$$

where  $1 \leq j \leq N$ ,  $\theta_h, \theta_j$  are phases of the hub and leaf nodes, respectively,  $K$  is the coupling strength, and  $\alpha$  is the phase shift. By introducing the phase difference  $\varphi_j = \theta_h - \theta_j$  and the natural frequency difference  $\Delta\omega_j = \omega_h - \omega_j$ , Eqs. (4.134) can be transformed to

$$\dot{\varphi}_j = \Delta\omega_j - K \sum_{j=1}^N \sin(\varphi_j + \alpha) - K \sin(\varphi_j - \alpha), j \in [1, N] \quad (4.135)$$

By introducing the mean-field order parameter

$$z(t) \equiv R(t)e^{i\psi(t)} = \frac{1}{N} \sum_{j=1}^N e^{i\varphi_j}, \quad (4.136)$$

Equations (4.135) can be rewritten as

$$\dot{\varphi}_j = fe^{i\varphi_j} + g + \bar{f}e^{-i\varphi_j}, \quad (4.137)$$

where  $j = 1, 2, \dots, N$ , and

$$f = iKe^{-i\alpha}/2, g = \Delta\omega - NKR \sin(\Psi + \alpha). \quad (4.138)$$

In terms of the OA ansatz, one can obtain the following equation of the order parameter:

$$\dot{z} = -\frac{K}{2}e^{-i\alpha}z^2 + i[\Delta\omega - NKR \sin(\Psi + \alpha)]z + \frac{K}{2}e^{i\alpha}. \quad (4.139)$$

The following procedure becomes easier in analyzing the collective dynamics of star-network coupled oscillators in terms of the order-parameter dynamics.

#### 4.4.6.2 Stationary Synchronous States

The above defined order parameter behaves in different ways as oscillators exhibit different collective dynamics. It can be noticed that only when  $R(t) = 1$  and the collective phase  $\Phi(t) = \text{constant}$ , a globally synchronous state on star network can be achieved. If the amplitude  $R(t) = 1$  while  $\Phi(t)$  is temporally periodic, then one has  $\varphi_j(t) = \varphi(t)$ , and this corresponds to the synchrony of leaf nodes while the hub are asynchronous to them. When  $R(t) = 0$ , oscillators behave incoherently, and  $0 < R(t) < 1$  corresponds to a partially synchronous state. On the other hand,  $R(t)$  could be time-independent or time-dependent, and the collective motion can be regular or chaotic, depending on system parameters and initial conditions.

By introducing  $z = x + iy$ , Eq. (4.139) is written as two real equations:

$$\begin{aligned} \dot{x} &= K_1(N+1/2)y^2 - K_1x^2/2 + K_2(N-1)xy - \Delta\omega y + K_1/2, \\ \dot{y} &= -K_2(N-1/2)x^2 - K_2y^2/2 - K_1(N+1)xy + \Delta\omega x + K_2/2. \end{aligned} \quad (4.140)$$

here  $K_1 = K \cos \alpha$ ,  $K_2 = K \sin \alpha$ . The fixed points  $(x_{1\sim 4}, y_{1\sim 4})$  can be worked out by setting  $\dot{x} = 0$ ,  $\dot{y} = 0$  as

$$x_{1,2} = \frac{(-\Delta\omega \pm A) \sin \alpha}{K(2N \cos 2\alpha + 1)}, \quad (4.141a)$$

$$y_{1,2} = \frac{(-\Delta\omega \pm A) \cos \alpha}{K(2N \cos 2\alpha + 1)}, \quad (4.141b)$$

$$x_{3,4} = \frac{\sin \alpha}{K} + \frac{[B \pm N(B - 2 \sin 2\alpha)] \cos \alpha}{K(N^2 + 2N \cos 2\alpha + 1)}, \quad (4.141c)$$

$$y_{3,4} = \frac{-\Delta\omega(-\cos \alpha \pm B \sin \alpha - K \cos \alpha)}{K(N^2 + 2N \cos 2\alpha + 1)}. \quad (4.141d)$$

The stability of these fixed points is summarized as follows:

The parameters in Table 4.1 can be analytically obtained as

$$K_c^f = \Delta\omega / \sqrt{2N \cos 2\alpha + 1}, \quad (4.142a)$$

$$K_{SC}^\pm = \mp \Delta\omega / (N \cos 2\alpha + 1), \quad (4.142b)$$

$$\alpha_0^\pm = \pm \arccos(-/N)/2 \quad (4.142c)$$

It is interesting that these fixed points correspond to different collective states in the star-networked systems:

**Table 4.1** Fixed points  $(x_{1\sim 4}, y_{1\sim 4})$  and their stable regions in the  $\alpha \sim K$  parameter space

Fixed Points	Stable regions
$(x_1, y_1)$	$K < K_c^f, \alpha \in (\alpha_0^-, 0)$
$(x_1, y_1)$	$K > 0, \alpha \in (-\pi/2, \alpha_0^-)$
$(x_2, y_2)$	$K > K_{SC}^+, \alpha \in (\alpha_0^+, \pi/2)$
$(x_3, y_3)$	$K > K_{SC}^-, \alpha \in (\alpha_0^-, 0)$
$(x_3, y_3)$	$K < K_{SC}^+, \alpha \in (\alpha_0^+, \pi/2)$
$(x_4, y_4)$	Always unstable

(1) **Synchronous State:**

Fixed points  $(x_{3,4}, y_{3,4})$  satisfy  $|z| = 1$ , which correspond to the **synchronous state (SS)** of the star, i.e. the phase difference between the hub and leaves keeps constant:

$$\varphi_j(t) = \text{const}, j = 1, 2, \dots, N. \quad (4.143)$$

It can be seen from Table 4.1 that  $(x_4, y_4)$  is always unstable, whilst the state  $(x_3, y_3)$  is stable in the region shown in Table 4.1.

(2) **Splay State:**

When the modulo of the fixed points  $(x_{1,2}, y_{1,2})$  satisfy  $|z| > 1$ , these points are unphysical. Only when  $|z| < 1$  the fixed points  $(x_{1,2}, y_{1,2})$  are related to the collective **splay state (SPS)**, i.e.

$$\varphi_j(t) = \varphi(t + jT/N), j = 1, 2, \dots, N, \quad (4.144)$$

where  $T$  is the period of  $\varphi(t)$ . This state indicates that leaf oscillators can achieve an ordered state with a fixed time delay between neighboring leaves. All possible stable regions of the SPS are given in Table 4.1.

**4.4.6.3 Periodic Synchronous States**

It should be emphasized that the long-term solutions of (4.140) include not only the stationary states  $(x_{1\sim 4}, y_{1\sim 4})$ , but also time-dependent states. Because the complex Eq. (4.139) is two-dimensional in real space, the possible time-dependent solution should be periodic. Further analysis indicates that the system possesses two types of periodic solutions, where type I exists in the parameter region  $0 < \alpha < \pi/2$  and  $K < K_{ec} = K_2$ , and the type-II periodic solution can be found for some special values of frustrations such as  $\alpha = 0, \pm\pi/2$ . These two types of periodic dynamics are different. We give a brief discussion of these two solutions.

(1) **In-phase State:**

Type-I periodic solution can be found by using the polar coordinate  $z = Re^{i\Psi}$ , and Eqs. (4.140) can be transformed to

$$\begin{aligned} \dot{R} &= -\frac{K}{2}(R^2 - 1) \cos(\Psi + \alpha), \\ \dot{\Psi} &= -\frac{K}{2}\left(R + \frac{1}{R}\right) \sin(\Psi - \alpha) + \Delta\omega - NK R \sin(\Psi + \alpha). \end{aligned} \quad (4.145)$$

Equations (4.145) have a *limit-cycle solution* with a radius  $R = 1$  and a cyclic evolution of the phase variable  $\Psi(t)$ . This periodic solution is related to the so-called **in-phase state (IPS)**, where phases of all oscillators are synchronous as.

$$\varphi_j(t) = \varphi(t), j = 1, 2, \dots, N, \tag{4.146}$$

i.e. all leaf oscillators are synchronous to each other while they are asynchronous to the hub oscillator. The stability of this limit cycle can be analyzed in terms of Floquet theory.

(2) **Neutral State:**

Type-II periodic solution occurs at some certain values of phase shifts. Let us take the case  $\alpha = 0$  as an example, where Eqs. (4.140) can be simplified as

$$\begin{aligned} \dot{x} &= K\left(\frac{1}{2} + N\right)y^2 - \frac{K}{2}x^2 - \Delta\omega y + \frac{K}{2} \\ \dot{y} &= -K(N + 1)xy + \Delta\omega x \end{aligned} \tag{4.147}$$

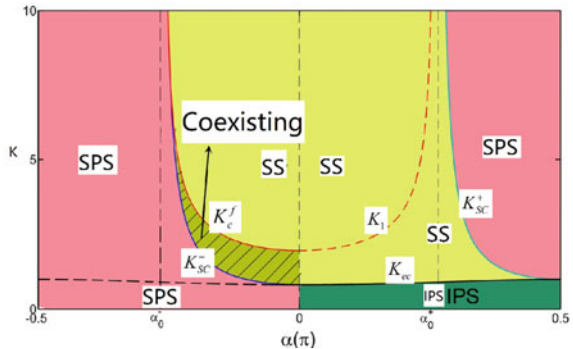
These two equations keep invariant under the time-reversal transformation.

$$\mathbf{R} : (t, x, y) \mapsto (-t, -x, y). \tag{4.148}$$

This leads to an attracting set in the phase plane  $x > 0$  and a repulsing set in the phase plane  $x < 0$ . If an orbit passes the boundary  $x = 0$  in both directions, this orbit will be a closed type and neutrally stable. Therefore when  $\alpha = 0$  Eqs. (4.145) have the neutral periodic solution, and this system is called a *quasi-Hamiltonian* system. The corresponding collective state is called the **neutral state** (NS). The NS depends on initial states, while the IPS is independent of initial states.

We further present the phase diagram in the  $\alpha \sim K$  space in Fig. 4.13, where the stable parameter regions of some typical collective states such as the SS, the SPS, and the IPS are shown in the phase diagram, respectively. The coexistence region of the incoherent state and the splay state is plotted by shadow. Three routes to synchronization are shown as the splay state to the synchronous state, the in-phase state to the synchronous state, and the neutral state to the synchronous state.

**Fig. 4.13** The phase diagram of the star-coupled oscillator system.  $\alpha$  is the phase shift, and  $K$  is the coupling strength. (Adapted from Ref. [108])



It can be seen from the above analysis of the two-dimensional order-parameter dynamics that it is much easier than the original  $(N + 1)$ -dimensional dynamics given by (4.134). Moreover, the dynamics of the order parameter can be completely solvable. Readers may refer [108–111] for a more detailed discussion.

## 4.5 Remarks

Collective behavior of a complex system implies the emergence of a macroscopic order from the organizations of populations of units with mutual interactions. Usually the number of degrees of freedom at the microscopic level is so large that an exact description of the system at this level is impossible and also unnecessary. Therefore a macroscopic study of the complex system is significant. In the description of thermodynamics of gas, one only needs to use several variables, e.g. the temperature, the pressure and the volume to depict the property of a large number of particles moving in a box. Generally, the macroscopic order of a complex system is originated from the reduction of degrees of freedom by transiting from microscopic to macroscopic levels, i.e. and the macroscopic description of the order needs only a few variables. These macroscopic variables are called the order parameters.

On the other hand, the emergence of order parameters in a system is spontaneous rather than manually selected. This means the ordered state is the result of self-organization and competitions of units. In this Chapter, we applied the slaving principle to discuss the competitions among various state variables with different time scales. We showed that at the critical point, only a few slow variables will conquer the large number of fast variables and dominate the evolution of the system, and these slow variables grow up to become the order parameters. In this case the dynamics can be described by the evolution of these small number of order parameters, which is a great dimension reduction.

It should be stressed that the idea of the reduction scheme can be extensively applied to various transitions, such as the emergence of spatiotemporal patterns, the transitions from partial to global synchronization in coupled oscillators, the transitions from non-integrability to integrability in Hamiltonian systems, thermodynamic phase transitions in condensed matters, and so on. We think the slaving principle and more generally the principles in synergetic theory and dissipative-structure theory are exhibiting their great privileges in exploring the order emergence of complex systems in recent years.

An effective reduction depends strongly on the motivation, and it can be a projection to a low-dimensional subspace, or a dimension reduction of high-dimensional systems, or a topology simplification, et al. The reducibility of complex systems should obey, at least approximately obey the following physical properties:

### (1) **Symmetry**

This is a well-known solvable case in physics, which applies also in the process of reduction. For example, symmetries in a Hamiltonian system lead to various

invariances such as invariant integrals and constants of motion, which are responsible for low-dimensional motions. Another example is the reduction problem in hydrodynamics, where the hydrodynamic collision invariance leads to conservation laws and furthermore the deduction of fluid dynamics such as Navier-Stokes equations and Euler equations.

(2) **Distinct time/space scales**

The slaving principle we discussed in this chapter is closely related to the separation of time scales of different degrees of freedom, which also provides a scheme in picking up order parameters from many state variables. This reminds us the emergence of order at the onset of phase transition in statistical physics, at which the slowing down effect can be observed and only a few stable modes become unstable and dominate the global behavior of the system.

Physically, the relaxation time scale is related to the correlation time scale, and the spatial diffusion scale is related to the spatial correlation scale. A large distinction time or space scale naturally leads to the separation of state variables. An impressive example is the theoretical foundation of Brownian motion, and the Langevin equation was proposed, which is a stochastic equation of motion including both deterministic and random forces due to the time-scale separation of the relaxational time and the rapid thermal fluctuation. Statistical dynamics of Brownian-related processes distributed from physics to chemistry and biology becomes an important subject.

The above criteria for a reducible system in many cases are constructive. The invariance and conservation related to symmetry can bring forth various ways in accomplishing the reduction procedure in terms of transformation invariance, invariant group, invariant manifold, and invariant subspace. Therefore the most important mission is to seek for these invariant elements and symmetries. The slaving principle can be facilitated in terms of adiabatic eliminations, time averaging, or the central-manifold theorem. In statistical physics, one often applies the projection operators to obtain the dynamics of a lower-dimensional distribution function.

In this Chapter we also applied the dimension-reduction schemes to study the emergence of sustained oscillation in networks of excitable units and gene-regulatory networks. Because collective rhythms are generated by the feedback mechanism, the topological motifs should be an important source of this feedback. For excitable networks, the feedback is organized by the loop structure (Winfree loop). One can apply the DPAD scheme to determine the phase order of units and reveal the embedded loops in a highly-complicated network. For gene-regulatory networks, the feedback is formed by an appropriate match of the active and repressive regulations among genes. A computation of functional weights can be applied to evaluate the dynamical contributions of topological networked links, where small functional-weight links can be eliminated by keeping dominant links. We should emphasize that these two techniques can be well applied to practical systems, where only data or time series are available to measure.

Synchronization of coupled oscillators is another vivid example of dynamical emergence, which covers a large extent of different topics. The ordered emergence is accompanied with a dimension reduction of microdynamics, so one may introduce

and explore the order-parameter dynamics in terms of self-consistency approach, the Ott-Antonsen ansatz or the ensemble order parameter technique. Symmetries exist in many networks of coupled phase-oscillator systems, Watanabe and Strogatz introduced the Mobius transformation (called the WS transformation) to reduce the high-dimensional dynamics to a three-dimensional space, and the two-dimensional invariant manifold obtained by using the OA ansatz is in fact a sub-manifold in the WS space. We will not discuss these topics here. Interested readers may refer [6, 92, 93] for more information.

## References

1. Needham, J.: Science and Civilisation in China, vol. 4. Science Press, Shanghai, Physics and related technologies (2010)
2. Laplace, P. S.: A philosophical essay on probabilities. Truscott, F. W., Emory, F. L., trans. New York: Wiley, (1902)
3. Anderson, P.W.: More is different: Broken symmetry and the nature of the hierarchical structure of science. *Science* **177**(4047), 393–396 (1972)
4. Palsson, B.O.: Systems Biology: Properties of Reconstructed Networks. Cambridge University Press, Cambridge (2006)
5. Meyers, R.A. (ed.): Encyclopedia of Complexity and Systems Science. Springer-Verlag, Berlin (2009)
6. Zheng, Z. G.: Emergence Dynamics in Complex Systems: From Synchronization to Collective Transport, vols. I and II, Science Press, Beijing (2019).
7. Prigogine, I., Nicolis, G.: Self-organization in Non-equilibrium Systems, from Dissipative Structures to Order Through Fluctuations. Wiley, New Jersey (1977)
8. Kauffman, S.A.: The Origins of Order: Self-Organization and Selection in Evolution. Oxford University Press, USA (1993)
9. Ott, E.: Chaos in Dynamical Systems. Cambridge University Press, Cambridge (1993)
10. Eckmann, J.P., Ruelle, D.: Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* **57**, 617 (1985)
11. Strogatz, S.H.: Nonlinear Dynamics and Chaos: With Applications to Physics, Chemistry, and Engineering. Westview Press;CRC Press, Biology (2018)
12. Fuchs, A.: Nonlinear Dynamics in Complex Systems: Theory and Applications for the Life-, Springer, Neuro- and Natural Sciences (2012)
13. Turing, A.M.: The chemical basis of morphogenesis. *Philos. Trans. R. Soc. B* **237**, 37–72 (1952)
14. Cross, M., Hohenberg, P.: Pattern formation outside of equilibrium. *Rev. Mod. Phys.* **65**, 851 (1993)
15. Cross, M., Greenside, H.: Pattern Formation and Dynamics in Nonequilibrium Systems. Cambridge University Press, Cambridge (2009)
16. Gromov, M., Capasso, V., Gromov, M., Harel-Bellan, A., Morozova, N., Pritchard, L.L. (eds.): Pattern Formation in Morphogenesis: Problems and Mathematical Issues. Springer-Verlag, Berlin Heidelberg (2013)
17. Huang, K.: Statistical Mechanics, 2nd edn. Wiley, New York (1987)
18. Stanley, H.E. (ed.): Introduction to Phase Transitions and Critical Phenomena. Oxford University Press, Oxford (1987)
19. Ma, S.K.: Modern theory of critical phenomena. Benjamin, New York (1976)
20. Haken, H.: Synergetics: An Introduction Nonequilibrium Phase Transitions and Self-Organization in Physics. Springer-Verlag, Berlin Heidelberg, Chemistry and Biology (1978)

21. Haken, H.: *Advanced Synergetics: Instability Hierarchies of Self-Organizing Systems and Devices*. Springer-Verlag, Berlin Heidelberg (1983)
22. Watts, D.J., Strogatz, S.H.: Collective dynamics of small-world networks. *Nature* **393**(6684), 440–442 (1998)
23. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. *Science* **286**, 509–512 (1999)
24. Albert, R., Barabási, A.L.: *Statistical mechanics of complex networks*. *Rev. Mod. Phys.* **74**(1), 47 (2002)
25. Eli Ben-Naim, H., Frauenfelder, Z., Torozckai Eds., *Complex Networks*, Lecture Notes in Physics, vol. 650, Springer, Amazon, (2004)
26. Dorogovtsev, S.N., Mendes, J.F.F.: *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press, Oxford (2003)
27. Pastor-Satorras, R., Vespignani, A.: *Evolution and Stricture of the Internet: A Statistical Physics Approach*. Cambridge University Press, Cambridge (2004)
28. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.U.: *Complex networks: Structure and dynamics*. *Phys. Rep.* **424**, 175–308 (2006)
29. Fang, J.Q., Wang, X.F., Zheng, Z.G., Bi, Q., Di, Z.R., Li, X.: New interdisciplinary science: network science (I). *Progress Phys* **27**(3), 239–343 (2007)
30. Fang, J.Q., Wang, X.F., Zheng, Z.G., Li, X., Di, Z.R., Bi, Q.: New interdisciplinary science: network science (II). *Progress Phys* **27**(4), 361–448 (2007)
31. Glass, L., Mackay, M.C.: *From Clocks to Chaos: The rhythms of Life*. Princeton University Press, Princeton, NJ (1988)
32. Zheng, Z.G.: Deciphering biological clocks, and reconstructing life rhythms. *Physics* **46**(12), 802–808 (2017)
33. The official website of the Nobel Prize: [https://www.nobelprize.org/nobel\\_prizes/medicine/laureates/2017/press.html](https://www.nobelprize.org/nobel_prizes/medicine/laureates/2017/press.html)
34. Mauro, Z., Rodolfo, C., Giuseppe, M., Chiaki, F., Gianluca, T.: Circadian Clocks: What Makes Them Tick? *Chronobiol. Int.* **17**(4), 433–451 (2000)
35. De Mairan.: *Observation botanique, Histoire de l'Académie royale dessciences avec les mémoires de mathématique et de physique tirés des registres decette Académie*, 35 (1729).
36. Konopka, R.J., Benzer, S.: Clock mutants of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **68**, 2112–2116 (1971)
37. Reddy, P., Zehring, W.A., Wheeler, D.A., Pirrotta, V., Hadfield, C., Hall, J.C., Rosbash, M.: Molecular analysis of the period locus in *Drosophila melanogaster* and identification of a transcript involved in biological rhythms. *Cell* **38**, 701–710 (1984)
38. Bargiello, T.A., Young, M.W.: Molecular genetics of a biological clock in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **81**, 2142–2146 (1984)
39. Bargiello, T.A., Jackson, F.R., Young, M.W.: Restoration of circadian behavioural rhythms by gene transfer in *Drosophila*. *Nature* **312**, 752–754 (1984)
40. Hardin, P.E., Hall, J.C., Rosbash, M.: Feedback of the *Drosophila* period gene product on circadian cycling of its messenger RNA levels. *Nature* **343**, 536–540 (1990)
41. Vitaterna, M.H., King, D.P., Chang, A.M., Kornhauser, J.M., Lowrey, P.L., McDonald, J.D., Dove, W.F., Pinto, L.H., Turek, F.W., Takahashi, J.S.: Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior. *Science* **264**(5159), 719–725 (1994)
42. Price, J.L., Blau, J., Rothenfluh, A., Abodeely, M., Kloss, B., Young, M.W.: double-time is a novel *Drosophila* clock gene that regulates PERIOD protein accumulation. *Cell* **94**, 83–95 (1998)
43. Nekorin, V. I.: *Introduction to Nonlinear Oscillations*, Wiley-VCH, (2015)
44. Guckenheimer, J., Holmes, P.: *Nonlinear Oscillations, Dynamical Systems, and Bifurcation of Vector Fields*. Springer-Verlag, New York (1983)
45. Glass, L.: Synchronization and Rhythmic Processes in Physiology. *Nature* **410**, 277–284 (2001)
46. Elowitz, M.B., Leibler, S.: A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000)



47. Buzsaki, G., Draguhn, A.: Neuronal oscillations in cortical networks. *Science* **304**, 1926–1929 (2004)
48. Zheng, Z. G.: *Collective Behaviors and Spatiotemporal Dynamics in Coupled Nonlinear Systems*. Beijing, Higher Education Press (in Chinese) (2004)
49. Goldbeter, A.: Computational approaches to cellular rhythms. *Nature* **420**, 238–245 (2002)
50. Ferrell, J.E., Tsai, T.Y., Yang, Q.: Modeling the cell cycle: Why do certain circuits oscillate? *Cell* **144**, 874–885 (2011)
51. Gardner, T.S., Cantor, C.R., Collins, J.J.: Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**(6767), 339–342 (2000)
52. Ozbudak, E., Thattai, M., Lim, H., et al.: Multistability in the lactose utilization network of *Escherichia coli*. *Nature* **427**, 737–740 (2004)
53. Tsai, T.Y., Choi, Y.S., Ma, W.Z., Pomerening, J.R., Tang, C., Ferrell, J.E.: Ferrell, robust, tunable biological oscillations from interlinked positive and negative feedback loops. *Science* **321**(5885), 126–129 (2008)
54. Huang, X.H., Zheng, Z.G., Hu, G., Wu, S., Rasch, M.J.: Different propagation speeds of recalled sequences in plastic spiking neural networks. *New J. Phys.* **17**, 035006 (2015)
55. Chen, L., Aihara, K.: Chaotic simulated annealing by a neural-network model with transient chaos. *Neur. Net.* **8**, 915–930 (1995)
56. Herz, A.V.M., Gollisch, T., Machens, C.K., Jaeger, D.: Modeling single-neuron dynamics and computations: A balance of detail and abstraction. *Science* **314**, 80–85 (2006)
57. Mandelblat-Cerf, Y., Novick, I., Vaadia, E.: Expressions of multiple neuronal dynamics during sensorimotor learning in the motor cortex of behaving monkeys, *PLoS ONE* **6**(7), e21626 (2011)
58. Li, N., Daie, K., Svoboda, K., Druckmann, S.: Robust neuronal dynamics in premotor cortex during motor planning. *Nature* **532**, 459–464 (2016)
59. Burke, J.F., Zaghoul, K.A., et al.: Synchronous and asynchronous theta and gamma activity during episodic memory formation. *J. Neur.* **33**, 292–304 (2013)
60. Qian, Y., Huang, X.D., Hu, G., Liao, X.H.: Structure and control of self-sustained target waves in excitable small-world networks. *Phys. Rev. E* **81**, 036101 (2010)
61. Qian, Y., Liao, X.H., Huang, X.D., Mi, Y.Y., Zhang, L.S., Hu, G.: Diverse self-sustained oscillatory patterns and their mechanisms in excitable small-world networks. *Phys. Rev. E* **82**, 026107 (2010)
62. Liao, X.H., Xia, Q.Z., Qian, Y., Zhang, L.S., Hu, G., Mi, Y.Y.: Pattern formation in oscillatory complex networks consisting of excitable nodes. *Phys. Rev. E* **83**, 056204 (2010)
63. Liao, X.H., Qian, Y., Mi, Y.Y., Xia, Q.Z., Huang, X.Q., Hu, G.: Oscillation sources and wave propagation paths in complex networks consisting of excitable nodes. *Front. Phys.* **6**, 124 (2011)
64. Mi, Y.Y., Zhang, L.S., Huang, X.D., Qian, Y., Hu, G., Liao, X.H.: Complex networks with large numbers of labelable attractors. *Europhys. Lett.* **95**, 58001 (2011)
65. Mi, Y.Y., Liao, X.H., Huang, X.H., Zhang, L.S., Gu, W.F., Hu, G., Wu, S.: Long-period rhythmic synchronous firing in a scale-free network. *PNAS* **25**, E4931–E4936 (2013)
66. Zhang, Z.Y., Ye, W.M., Qian, Y., Zheng, Z.G., Huang, X.H., Hu, G.: Chaotic motifs in gene regulatory networks. *PLoS ONE* **7**, e39355 (2012)
67. Zhang, Z.Y., Li, Z.Y., Hu, G., Zheng, Z.G.: Exploring cores and skeletons in oscillatory gene regulatory networks by a functional weight approach. *Europhys. Lett.* **105**, 18003 (2014)
68. Zhang, Z.Y., Huang, X.H., Zheng, Z.G., Hu, G.: Exploring cores and skeletons in oscillatory gene regulatory networks by a functional weight approach. *Sci. Sin.* **44** 1319 (in Chinese) (2014)
69. Jahnke, W., Winfree, A.T.: A survey of spiral wave behavior in the oregonator model. *Int. J. Bif. Chaos* **1**, 445–466 (1991)
70. Courtemanche, M., Glass, L., Keener, J.P.: Instabilities of a propagating pulse in a ring of excitable media. *Phys. Rev. Lett.* **70**, 2182–2185 (1993)
71. Qian, Y., Cui, X.H., Zheng, Z.G.: Minimum Winfree loop determines self-sustained oscillations in excitable Erdos-Renyi random networks. *Sci. Rep.* **7**, 5746 (2017)

72. Zheng, Z.G., Qian, Y.: Self-sustained oscillations in biological excitable media. *Chin. Phys. B* **27**(1), 018901 (2018)
73. Qian, Y., Zhang, G., Wang, Y.F., Yao, C.G., Zheng, Z.G.: Winfree loop sustained oscillation in two-dimensional excitable lattices: Prediction and Realization. *Chaos* **29**, 073106 (2019)
74. Qian, Y., Wang, Y., Zhang, G., Liu, F., Zheng, Z.G.: Collective sustained oscillations in excitable small-world networks: the moderate fundamental loop or the minimum Winfree loop? *Nonlinear Dyn.* **99**(2), 1415 (2020)
75. Zhang, Z.Y., Zheng, Z.G., Niu, H.J., Mi, Y.Y., Wu, S., Hu, G.: Solving the inverse problem of noise-driven dynamic networks. *Phys. Rev. E* **91**, 012814 (2015)
76. Chen, Y., Wang, S.H., Zheng, Z.G., Zhang, Z.Y., Hu, G.: Depicting network structures from variable data produced by unknown colored-noise driven dynamics. *Europhys. Lett.* **113**, 18005 (2016)
77. Pikovsky, A., Rosenblum, M., Kurths, J.: *Synchronization, A Universal Concept in Nonlinear Sciences*. Cambridge University Press, New York (2001)
78. Strogatz, S.: *Sync: The emerging science of spontaneous order*, Hyperion, (2003)
79. Smith, H.M.: Synchronous flashing of fireflies. *Science* **82**, 151 (1935)
80. Buck, J.B.: Synchronous rhythmic flashing of fireflies. *Quart. Rev. Biol.* **13**(3), 301–314 (1938); Synchronous rhythmic flashing of fireflies. II. *Quart. Rev. Biol.* **63**(3), 265–289 (1988)
81. Huygenii, G.: *Horoloquim Oscillatorium Parisiis*, France, (1673)
82. Winfree, A.T.: Biological rhythms and the behavior of populations of coupled oscillators. *J. Theo. Biol.* **16**, 15–42 (1967)
83. Kuramoto, Y.: Self-entrainment of a population of coupled non-linear oscillators. In: *International Symposium on Mathematical Problems in Theoretical Physics*, Springer, Berlin/Heidelberg, (1975), NBR 6023
84. Winfree, A.T.: *Geometry of Biological Time*. Springer-Verlag, New York (1990)
85. Kuramoto, Y.: *Chemical Oscillations, Waves and Turbulence*. Springer-Verlag, Berlin (1984)
86. Acebrón, J.A., Bonilla, L.L., Vicente, C.J.P., Ritort, F., Spigler, R.: The Kuramoto model: A simple paradigm for synchronization phenomena. *Rev. Mod. Phys.* **77**(1), 137 (2005)
87. Arenas, A., Díaz-Guilera, A., Kurths, J., Moreno, Y., Zhou, C.: Synchronization in complex networks. *Phys. Rep.* **469**(3), 93–153 (2008)
88. Rodrigues, F.A., Peron, T.K.D.M., Ji, P., Kurths, J.: The Kuramoto model in complex networks, *Phys. Rep.* **610**, 1–98 (2016)
89. Osipov, G.V., Kurths, J., Zhou, C.S.: *Synchronization in Oscillatory Networks*, Springer Series in Synergetics, Springer-Verlag, Berlin, (2007)
90. Yao, N., Zheng, Z.G.: Chimera states in spatiotemporal systems: Theory and applications. *Int. J. Mod. Phys. B* **30**(7), 1630002 (2016)
91. Ott, E., Antonsen, T.M.: Low dimensional behavior of large systems of globally coupled oscillators. *Chaos* **18**(3), 037113 (2008)
92. Ott, E., Antonsen, T.M.: Long time evolution of phase oscillator systems. *Chaos* **19**(2), 023117 (2009)
93. Marvel, S.A., Strogatz, S.H.: Invariant submanifold for series arrays of Josephson junctions. *Chaos* **19**(1), 013132 (2009)
94. Marvel, S.A., Mirollo, R.E., Strogatz, S.H.: Identical phase oscillators with global sinusoidal coupling evolve by mobius group action. *Chaos* **19**(4), 043104 (2009)
95. Hu, G., Xiao, J.H., Zheng, Z.G.: *Chaos Control*. Shanghai Sci. Tech. Edu. Pub. House, Shanghai (2000)
96. Zheng, Z.G.: *Collective Behaviors and Spatiotemporal Dynamics in Coupled Nonlinear Systems*. Higher Education Press, Beijing (2004)
97. Zheng, Z.G., Hu, G., Hu, B.: Phase slips and phase synchronization of coupled oscillators. *Phys. Rev. Lett.* **81**, 5318–5321 (1998)
98. Zheng, Z.G., Hu, B., Hu, G.: Collective phase slips and phase synchronizations in coupled oscillator systems. *Phys. Rev. E* **62**, 402–408 (2000)
99. Hu, B., Zheng, Z.G.: Phase synchronizations: transitions from high- to low-dimensional tori through chaos. *Inter. J. Bif. Chaos* **10**(10), 2399–2414 (2000)

100. Zheng, Z.G.: Synchronization of coupled phase oscillators, Chap. 9. In: *Advances in Electrical Engineering Research*. vol. 1, pp: 293–327. (Ed.: Brouwer, T.M.) Nova Sci., (2011)
101. Gao, J., Xu, C., Sun, Y., Zheng, Z.G.: Order parameter analysis for low-dimensional behaviors of coupled phase-oscillators. *Sci. Rep.* **6**, 30184 (2016)
102. Zheng, Z.G., Zhai, Y.: Chimera state: From complex networks to spatiotemporal patterns (in Chinese). *Sci. Sin-Phys. Mech. Astron* **50**, 010505 (2020)
103. Xu, C., Xiang, H., Gao, J., Zheng, Z.G.: Collective dynamics of identical phase oscillators with high-order coupling. *Sci. Rep.* **6**, 31133 (2016)
104. Xu, C., Boccaletti, S., Guan, S.G., Zheng, Z.G.: Origin of bellerophon states in globally coupled phase oscillators. *Phys. Rev. E* **98**, 050202(R) (2018)
105. Xu, C., Gao, J., Xiang, H., Jia, W., Guan, S., Zheng, Z.G.: Dynamics of phase oscillators in the Kuramoto model with generalized frequency-weighted coupling. *Phys. Rev. E* **94**, 062204 (2016)
106. Xu, C., Sun, Y.T., Gao, J., Qiu, T., Zheng, Z.G., Guan, S.G.: Synchronization of phase oscillators with frequency-weighted coupling. *Sci. Rep.* **6**, 21926
107. Xu, C., Boccaletti, S., Zheng, Z.G., Guan, S.G.: Universal phase transitions to synchronization in Kuramoto-like models with heterogeneous coupling. *New J. Phys.* **21**, 113018 (2019)
108. Xu, C., Zheng, Z.G.: Bifurcation of the collective oscillatory state in phase oscillators with heterogeneity coupling. *Nonlinear Dyn.* **98**(3), 2365–2373 (2019)
109. Xu, C., Gao, J., Sun, Y.T., Huang, X., Zheng, Z.G.: Explosive or Continuous: Incoherent state determines the route to synchronization. *Sci. Rep.* **5**, 12039 (2015)
110. Chen, H.B., Sun, Y.T., Gao, J., Xu, C., Zheng, Z.G.: Order parameter analysis of synchronization on star networks. *Front. Phys.* **12**(6), 120504 (2017)
111. Xu, C., Sun, Y.T., Gao, J., Jia, W.J., Zheng, Z.G.: Phase transition in coupled star network. *Nonlinear Dyn.* **94**, 1267–1275 (2018)
112. Xu, C., Gao, J., Boccaletti, S., Zheng, Z.G., Guan, S.G.: Synchronization in starlike networks of phase oscillators. *Phys. Rev. E* **100**, 012212 (2019)
113. Gómez-Gardeñes, J., Gómez, S., Arenas, A., Moreno, Y.: Explosive synchronization transitions in scale-free networks. *Phys. Rev. Lett.* **106**, 128701 (2011)
114. Ji, P., Peron, T.K.D., Menck, P.J., Rodrigues, F.A., Kurths, J.: Cluster explosive synchronization in complex networks. *Phys. Rev. Lett.* **110**, 218701 (2013)
115. Leyva, I., et al.: Explosive first-order transition to synchrony in networked chaotic oscillators. *Phys. Rev. Lett.* **108**, 168702 (2012)
116. Zou, Y., Pereira, T., Small, M., Liu, Z., Kurths, J.: Basin of attraction determines hysteresis in explosive synchronization. *Phys. Rev. Lett.* **112**, 114102 (2014)
117. Rodrigues, F.A., Peron, T.K.D., Ji, P., et al.: The Kuramoto model in complex networks. *Phys. Rep.* **610**, 1–98 (2016)

# Chapter 5

## Basics of Molecular Modeling and Molecular Simulation



Chenyu Tang (唐晨宇)  and Yanting Wang (王延颢) 

**Abstract** Molecular simulation is a powerful tool in computational physics, with which researchers are enabled to study physical properties of molecular systems at the atomic level. It is therefore of much vitality and one of the most fundamental and effective ways for researchers to study soft-matter physics and biophysics. This chapter majorly focuses on introducing the principles and some applications of molecular modeling and molecular simulation, including Monte Carlo simulation and molecular dynamics simulation. It also presents a synopsis to first-principle calculations, as well as some basic concepts of equilibrium statistical physics that many researchers may find useful. The chapter is collaboratively written by Chenyu Tang (OrcID: 0000-0002-6914-7348) and Prof. Yanting Wang (OrcID: 0000-0002-0474-4790).

**Keywords** Molecular modeling · Monte Carlo simulation · Molecular dynamics simulation

### 5.1 Molecular Simulation

Molecular simulation is an important tool in computational physics, which can simulate materials systems at the molecular level. It can not only calculate microscopic properties of simulated systems, but also predict their macroscopic properties via the support of statistical physics.

---

C. Tang (唐晨宇) · Y. Wang (王延颢) (✉)  
Institute of Theoretical Physics, Chinese Academy of Sciences,  
55 East Zhongguancun Road, P.O. Box 2735, Beijing 100190, China  
e-mail: [wangyt@itp.ac.cn](mailto:wangyt@itp.ac.cn)

### ***5.1.1 Computational Physics: A Bridge Connecting Theories and Experiments***

Since the beginning of the usage of computers in science, computational physics has become an important branch of physics study. Before the appearance of computational physics, approximations usually had to be taken when theoretical studies were conducted to investigate physical problems. Such approximations often came with inevitable disadvantages that most of the time the consequence of making approximations could not be estimated ahead of time. On the other hand, experimental results are usually too comprehensive to establish the causality among various factors. Therefore, it was a common practice that a large gap between theory and experiment existed when investigating a certain physical system.

The emergence of computational physics provides contemporary researchers a brand new method in predicting the properties of many systems in ways that have never been done before. It fills in the gap between theory and experiment by generating accurate (in the sense that the underlying model is good) numerical results while keeping input variables controllable. Current researches have been following a certain route in terms of using computational method. The theoretical works often offer a model to specific systems, which renders certain directions for computations to follow, whereas the results of computational works feed theories with solutions that may not be easily found without numerical calculations. The computations further play a significant role in advising experimentalists in conducting related experiments, whose work in return marks parameters for computational physics, which are essential in narrowing the gap between simulated systems and the real ones.

There are certain typical functions of computational physics used under most circumstances among studies. One of them is to numerically solve analytical equations. In some cases, it triggers much difficulty to solve equations with analytical approach. It is thus unwise and sometimes impossible to solve most sets of equations, no matter differential or integral ones, without numerical calculations. Bringing in computational methods as an effective way to provide sets of numerical solutions are of much necessity. Another major function resides in the simulation of many-body problems to obtain more realistic results. The many-body systems usually require sets of equations with too many degrees of freedom for physicists and mathematicians to draw analytical solutions. Although their approximations, often using perturbation or series expansion methods, have been remarkably successful in depicting simple systems, they lead to unacceptable deviations in most classical or quantum many-body systems. In avoiding such deviations, it is wise to use computational methods that offer much more precise results with iterative algorithms in due systems.

Moreover, there are certain systems that mere experiments may be insufficient to describe. Some of the experiments should be conducted under extreme conditions that is of overwhelming difficulty in reality. Computational physics therefore becomes crucially needed because manipulating simulated systems renders much more freedom than doing so under real laboratorial environments. Such freedom

is also manifested in creating systems to investigate theories using many approximations and ideal conditions, such as the Ising model. The usage of simulated computational methods allows perceptual understanding for the properties under these circumstances, leading to better comprehension for the fundamental concepts in theoretical physics.

To sum up, computational physics is gaining its importance with its unique and irreplaceable role in connecting theoretical and experimental physics. Its major functions may be concluded into four basic aspects: (1) solving analytical equations numerically; (2) simulating many-body systems to obtain more realistic results; (3) simulating under extreme conditions; and (4) simulating systems with many approximations or under ideal conditions. With the development of information technology that dramatically increases the power of parallel computing, computational physics will certainly have a rosy prospect in upcoming physics researches. It is thus vital for future researchers to acknowledge its significance and learn to better understand as well as make better usage of it in their respective studies.

### ***5.1.2 Molecular Simulations: Studying Physical Properties at the Molecular Level***

As described above, computational physics basically includes two aspects: providing numerical solutions to analytical equations and conducting computer simulation. To study fundamental principles of statistical physics and to understand certain physical properties of materials, computer simulations are of much significance. All materials simulations are roughly conducted at three temporal and spatial levels: macroscopic, mesoscopic, and microscopic. It is vital to make such a division for computer simulations in order to determine the correct parameters in use, for the systems at different levels vary dramatically in terms of their mechanical, thermodynamic, and electromagnetic properties. In this chapter, we will focus mainly on molecular simulations dealing with materials at the microscopic level with the resolution of atoms, the mechanisms of different types of molecular simulations, and their applications in soft-matter physics.

Types of molecular simulations with especial significance in computer simulations are Monte Carlo (MC) simulations and molecular dynamics (MD) simulations. The inclusion of such simulations provides researchers with two fundamentally different measures in dealing with different problems, the differences and usages of which will also be further demonstrated and introduced in the following parts of this chapter.

A basic knowledge that most researchers should have acquired when investigating into computer simulation is the vitality of importance sampling for the Boltzmann distribution, which to a large extent signifies that the basis of molecular simulation resides in statistical physics. Thus, a good understanding of statistical physics is vital for researchers to utilize molecular simulations efficiently in their scientific investigations. On the other hand, a good molecular model is prerequisite for molecular

simulations. An empirical classical molecular model can be built up upon data from experiments and/or first-principles calculations. Therefore, we will start with a brief overview of the first-principles calculations.

## 5.2 First-Principles Calculations

First-principles calculations compute physical and chemical properties of molecular systems with the resolution of nuclei and electrons by numerically solving quantum mechanics equations. The purpose of this introduction is to show how the interactions between atoms at zero temperature of a non-relativistic time-independent molecular system can be accurately calculated at the quantum mechanics level, whose results can be used, sometimes along with experimental data, to build up empirical classical molecular models.

### 5.2.1 Electronic Structure Methods

For a non-relativistic time-independent molecular system, the aim of first-principles calculations is to ultimately solve the many-body Schrödinger equation:

$$\hat{H}_{\text{tot}} \left| \Psi_{\text{tot}} \left( \left\{ \vec{r}_n, \vec{r}_e \right\} \right) \right\rangle = E_{\text{tot}} \left| \Psi_{\text{tot}} \left( \left\{ \vec{r}_n, \vec{r}_e \right\} \right) \right\rangle \quad (5.2.1)$$

where both the Hamiltonian  $\hat{H}_{\text{tot}}$  and wave function  $\Psi_{\text{tot}}$  are dedicated for all nuclei and electrons. The subscript n denotes nuclei and e denotes electron, and  $\vec{r}$  denotes the space degrees of freedom. The total Hamiltonian operator can be separated into several parts:

$$\hat{H}_{\text{tot}} = \hat{T}_n + \hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee} + \hat{V}_{nn} \quad (5.2.2)$$

where  $\hat{T}$  represents the kinetic energy operator and  $\hat{V}$  the potential energy.

The above Schrödinger equation has no analytical solution except for extremely simple systems such as a hydrogen atom. In the following of this part, we are going to introduce a few well-developed methods for numerically solving this many-body Schrödinger equation for molecular systems by applying some approximations. Those methods can be classified into two types: electronic structure methods based on the Hartree-Fock (HF) approximation and density functional theory (DFT) methods based on the DFT developed by Walter Kohn et al.

### 5.2.1.1 Born-Oppenheimer Approximation

Almost all first-principles calculations aimed to solve the many-body Schrödinger equation have to begin with the Born-Oppenheimer (BO) approximation which divides the calculations of nuclei and electron interactions into two separate steps.

1. Fix nuclei and solve the wave functions related to electrons. Because nuclei move much slower than electrons, it is suitable to assume that the fast-moving electrons evolve while the nuclei are stationary. Accordingly, the total Hamiltonian can be expressed into two terms:

$$\hat{H}_{\text{tot}} = \hat{T}_{\text{n}} + \hat{H}_{\text{e}} \quad (5.2.3)$$

where

$$\hat{H}_{\text{e}} = \hat{T}_{\text{e}} + \hat{V}_{\text{ne}} + \hat{V}_{\text{ee}} + \hat{V}_{\text{nn}} \quad (5.2.4)$$

$$\hat{T}_{\text{e}} = -\sum_{i=1}^N \frac{1}{2} \nabla_i^2, \quad \hat{V}_{\text{ne}} = -\sum_{i=1}^N \sum_{a=1}^{N_a} \frac{Z_a}{|\vec{r}_{na} - \vec{r}_{ei}|},$$

$$\hat{V}_{\text{ee}} = \sum_{i=1}^N \sum_{j>i}^N \frac{1}{|\vec{r}_{ei} - \vec{r}_{ej}|}, \quad \hat{V}_{\text{nn}} = \sum_{a=1}^{N_a} \sum_{b>a}^{N_b} \frac{Z_a Z_b}{|\vec{r}_{na} - \vec{r}_{nb}|}$$

with  $Z_a$  the charge number of nucleus  $a$ . The total wave function can be approximated by the uncorrelated contributions from nuclei and electrons:

$$\Psi_{\text{tot}}(\{\vec{r}_{\text{n}}, \vec{r}_{\text{e}}\}) \approx \Psi_{\text{n}}(\{\vec{r}_{\text{n}}\}) \Psi_{\text{e}}(\{\vec{r}_{\text{e}}\}) \quad (5.2.5)$$

2. Fix obtained electron wave functions and update nuclei positions. Suppose the above electron part  $E = \langle \Psi_{\text{e}} | \hat{H}_{\text{e}} | \Psi_{\text{e}} \rangle$  is successfully solved, we can then update the nuclei degrees of freedom by fixing electron wave functions and solving  $(\hat{T}_{\text{n}} + E)|\Psi_{\text{n}}\rangle = E_{\text{tot}}|\Psi_{\text{n}}\rangle$  accordingly.

The second step is easy, so the major task of a regular first-principles calculation is to solve the electron wave function in the first step. The BO approximation leads to a negligible relative error of less than  $10^{-4}$  except for extremely light atoms, such as hydrogen, in very special cases. Therefore, the BO approximation can be safely applied for almost all cases.

### 5.2.1.2 Hartree-Fock Approximation

In dealing with the wave functions for electrons, the correlations between electrons are first neglected and only the antisymmetric and indistinguishable properties of



electrons as Fermions are considered, so that the total electron wave function can be expressed as a linear combination of the single-electron wave functions of all electrons. This approximation allows the electron wave functions of electrons to be expressed as the so-called *Slater determinant*:

$$\Psi_e\left(\left\{\vec{r}_e\right\}\right) \approx |\Phi\rangle = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(x_1) & \phi_2(x_1) & \dots & \phi_N(x_1) \\ \phi_1(x_2) & \phi_2(x_2) & \dots & \phi_N(x_2) \\ \dots & \dots & \dots & \dots \\ \phi_1(x_N) & \phi_2(x_N) & \dots & \phi_N(x_N) \end{vmatrix} \quad (5.2.6)$$

where  $\phi_i$  is the single-electron wave function of electron  $i$ , and  $x$  include both the space and spin degrees of freedom. This simple ansatz for the wave function  $\Phi$  captures much of the physics required for accurate solutions of the Hamiltonian. Most importantly, the wave function is antisymmetric with respect to an interchange of any two electron positions, which is required by the Pauli Exclusion Principle:

$$\Phi(x_1, x_2 \dots x_i \dots x_j \dots x_N) = -\Phi(x_1, x_2 \dots x_j \dots x_i \dots x_N) \quad (5.2.7)$$

This wave function can be inserted into the Hamiltonian and the system energy can be rewritten as:

$$\begin{aligned} E &= \sum_{i=1}^N h_i + \sum_{i=1}^N \sum_{j>i}^N (J_{ij} - K_{ij}) + V_{nn} \\ &= \sum_{i=1}^N h_i + \frac{1}{2} \sum_{\substack{i,j=1 \\ i \neq j}}^N (J_{ij} - K_{ij}) + V_{nn} \end{aligned} \quad (5.2.8)$$

where

$$\begin{aligned} \hat{J}_j |\phi_i(i)\rangle &= \langle \phi_j(j) | \hat{g}_{ij} | \phi_j(j) \rangle |\phi_i(i)\rangle \\ \hat{K}_j |\phi_i(i)\rangle &= \langle \phi_j(j) | \hat{g}_{ij} | \phi_i(j) \rangle |\phi_j(i)\rangle \end{aligned} \quad (5.2.9)$$

Knowing that we may choose  $\phi$  to be an orthonormal set, we can introduce the Lagrange Multiplier  $\varepsilon_i$  to impose the condition that  $\phi$  are normalized, and minimize with respect to  $\phi$ :

$$\frac{\delta}{\delta \phi} \left[ \langle \hat{H} \rangle - \sum_j \varepsilon_j \int |\phi_j|^2 d\mathbf{r} \right] = 0 \quad (5.2.10)$$

which reduces to a set of single-electron equations of the form:

$$-\frac{1}{2}\nabla^2\phi_i(\mathbf{r}) + V_{\text{ion}}(\mathbf{r})\phi_i(\mathbf{r}) + U(\mathbf{r})\phi_i(\mathbf{r}) = \varepsilon_i\phi_i(\mathbf{r}) \quad (5.2.11)$$

We can obtain a quasi-eigen equation, which is called the HF equation:

$$\begin{aligned} \varepsilon_i|\phi_i\rangle &= \left(-\frac{1}{2}\nabla^2 + V_{\text{ion}}(\mathbf{r})\right)\phi_i(\mathbf{r}) + \sum_j \int d\mathbf{r}' \frac{|\phi_j'(\mathbf{r}')|^2}{|\mathbf{r}-\mathbf{r}'|} \phi_i(\mathbf{r}) - \sum_j \delta\sigma_i\sigma_j \int d\mathbf{r}' \frac{\phi_j^*(\mathbf{r}')\phi_i(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \phi_j(\mathbf{r}) \\ &= \hat{F}_i|\phi_i\rangle \end{aligned} \quad (5.2.12)$$

where

$$\hat{F}_i = \hat{h}_i + \sum_{\substack{j=1 \\ j \neq i}}^N (\hat{J}_j - \hat{K}_j) \quad (5.2.13)$$

The Fock operator  $\hat{F}_i$  in the above equations consists of three terms. Because the Fock operator depends on the solution of wave functions, the HF equation is a quasi-eigen equation which should be solved self-consistently. The term  $\hat{h}_i$  consists of the kinetic energy contribution and the electron-ion potential. The term  $\sum_{\substack{j=1 \\ j \neq i}}^N \hat{J}_j$ , or Hartree term, is simply the electrostatic potential arising from the charge distribution of  $N$  electrons. The exchange term  $\sum_{\substack{j=1 \\ j \neq i}}^N \hat{K}_j$  results from our inclusion of the Pauli principle and the assumed determinant form of the wave function, which has no classic counterpart.

The SCF technique is used to solve the HF equation self-consistently, which expands the wave functions by an assigned basis set,

$$|\phi_i'\rangle = \sum_{\alpha=1}^M c_{\alpha i} |\chi_{\alpha}\rangle \quad (5.2.14)$$

which leads to the Roothan-Hall equation in the matrix form:

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\varepsilon} \quad (5.2.15)$$

where  $\mathbf{F}$  is the Fock matrix,  $\mathbf{C}$  is the coefficient matrix,  $\mathbf{S}$  is the overlap matrix of the basis functions, and  $\boldsymbol{\varepsilon}$  is the matrix of orbital energies. Accordingly, it should be solved iteratively since  $\mathbf{F}$  and  $\mathbf{C}$  depend on each other.

Assuming a single-determinant form for the wave function, the HF theory neglects correlation between electrons. As a result, the electrons are subject to an *average* non-local potential arising from the other electrons, which leads to about 2% deviation of the total energy. Although the deviation sounds small, most of it comes from the outmost electrons which determines the physical and chemical properties of the molecular system. Thus, it is usually insufficient to compute at the HF level and the correlation effect should be included in some ways.

### 5.2.1.3 Electron Correlation Methods

On top of the HF method, the electron correlation methods incorporate the correction for electron correlations into the HF equations to obtain more accurate computational results. To achieve such corrections, several different methods can be applied, which will be briefly introduced as follows.

#### Configuration Interaction (CI) Method

The basis for the CI method is very simple that an *exact* many-body wave function may be written as a linear combination of the Slater determinants  $\Phi_i$ :

$$\Psi = a_0\Phi_{\text{HF}} + \sum_{i=1} a_i\Phi_i \quad (5.2.16)$$

where  $\Phi_{\text{HF}}$  is the Hartree-Fock determinant,  $\Phi_i$  is the determinant with some electrons virtually “excited” from the ground state in  $\Phi_{\text{HF}}$ , and  $a_0$  and  $a_i$  are parameters. The fundamental issue for the CI method is to obtain a satisfying accuracy while keeping the expansion length as short as possible. It can be proven that the single excitation leads to zero, so the least accurate expansion is CISD, where S refers to singlet and D refers to doublet. There are more accurately CISDT and CISDTQ, where T refers to triplet and Q refers to quadruplet. Leaving more terms in the expression results in a better accuracy of the CI method.

#### Many-Body Perturbation (MP) Method

The many-body perturbation method, based on the quantum perturbation theory, is almost synonymous to the Møller-Plesset method, its most popular implementation, which treats electron correlation as a perturbation to the Hartree-Fock wave function. The accordance of its computational complexity and the correction to the correlation energy is listed below, where  $M$  is the size of the basis set.

Method	Correlation energy	Computational complexity
MP2	~80 to 90%	$O(M^5)$
MP3	~90 to 95%	$O(M^6)$
MP4	~95 to 98%	$O(M^7)$

### Coupled Cluster (CC) Method

The coupled-cluster (CC) method resolves the problem of size extensibility, and is often very accurate, but more expensive than (limited) CI. The CC method assumes an exponential ansatz for the wave function

$$|\Psi_{CC}\rangle = e^{\hat{T}}|\Phi_0\rangle \quad (5.2.17)$$

where

$$e^{\hat{T}} = 1 + \hat{T} + \frac{1}{2}\hat{T}^2 + \frac{1}{6}\hat{T}^3 + \dots = \sum_{k=0}^{\infty} \frac{1}{k!}\hat{T}^k \quad (5.2.18)$$

The approximate computational complexity of CI methods, MP methods, and CC methods can be estimated as the chart given below.

Computational complexity	CI	MP	CC
M <sup>5</sup>		MP2	
M <sup>6</sup>	CISD	MP3, MP4(SDQ)	CCSD
M <sup>7</sup>		MP4	CCSD(T)
M <sup>8</sup>	CISDT	MP5	CCSDT
M <sup>9</sup>		MP6	
M <sup>10</sup>	CISDTQ	MP7	CCSDTQ

In practice, the user has to choose an appropriate method with a good balance between accuracy and complexity which well fits the designated problem. For instance, the HF method is generally used to locally optimize a configuration, the MP2 method is widely used to obtain a rough result with acceptable accuracy and computational cost, and the CC series are usually used to obtain very accurate results for the purpose of calibration.

The electronic structure methods are slower than the DFT methods described below, but have the benefit that higher level calculations guarantee to have a better accuracy.

### 5.2.2 Density Functional Theory

In dealing with many-body Schrödinger Equations, DFT employs electronic density distribution to replace electronic positions to be the variables of the wave function, which to a large extent simplifies the computational complexity and enables this method to be widely used in solving such equations. To use DFT methods in solving many-body Schrödinger Equations with the form of Eq. (2.1), the key is to replace

the electron positions  $\mathbf{r}$  by the electron density  $\rho(\{\mathbf{r}\})$ , and thus the wave function becomes a functional of the electron density. With that, Eq. (2.1) can be written in the form:

$$\hat{H}\Psi(\rho(\{\mathbf{r}\})) = E\Psi(\rho(\{\mathbf{r}\})) \quad (5.2.19)$$

Two theorems described below are the basis of solving the above equation numerically.

### 5.2.2.1 The Hohenberg-Kohn Theorem

The Hohenberg-Kohn Theorem [1] gives a statement that if there are  $N$  interacting electrons moving in an external potential  $V_{\text{ext}}(\mathbf{r})$ , the ground-state electron density  $\rho_0(\mathbf{r})$  minimizes the functional

$$E[\rho] = F[\rho] + \int \rho(\mathbf{r})V_{\text{ext}}(\mathbf{r})d\mathbf{r} \quad (5.2.20)$$

where  $F[\rho]$  is a functional of  $\rho$ . The minimum of  $E[\rho]$  could be observed in the ground-state, where it essentially becomes  $E_0$ . This can be proved simply in M. Levy's work [2], which will not be given in this section. Interested readers can check his work to better understand this statement.

The Hohenberg-Kohn Theorem implicates two basics that can be applied by DFT. The first is that it shows that all physical properties can be determined by electron density and it is a one-to-one relation between system energy and electron density. Secondly, system energy determined by an arbitrary electron density is always larger or equal to the real system energy, which is also referred to as the *variational principle for electron density*.

### 5.2.2.2 The Kohn-Sham Equations

To solve many-body Schrödinger Equation with electron density, it is vital to apply the Lagrange Undetermined Multiplier and the Variational Principle for electron density. Kohn and Sham [3] developed a set of differential equations to find the ground-state  $\rho_0(\mathbf{r})$  by separating  $F[\rho]$  into three terms, and the total energy Eq. (2.20) becomes:

$$E_{\text{DFT}}[\rho] = T_{\text{S}}[\rho] + E_{\text{ne}}[\rho] + E_{\text{XC}}[\rho] + J[\rho] \quad (5.2.21)$$

where

$$J[\rho] = \int \rho(\mathbf{r})V_{\text{ext}}(\mathbf{r})d\mathbf{r} \quad (5.2.22)$$

$$E_{\text{ne}}[\rho] = \frac{1}{2} \int \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' \quad (5.2.23)$$

$T_{\text{S}}[\rho]$  is the term representing kinetic energy of a non-interacting electron gas, not the total kinetic energy of the system,

$$T_{\text{S}}[\rho] = -\frac{1}{2} \sum_{i=1}^N \int \psi_i^*(\mathbf{r}) \nabla^2 \psi_i(\mathbf{r}) d\mathbf{r} \quad (5.2.24)$$

The term  $E_{\text{XC}}[\rho]$  represents the unknown exchange-correlation energy:

$$E_{\text{XC}}[\rho] = (T[\rho] - T_{\text{S}}[\rho]) + K[\rho] \quad (5.2.25)$$

where

$$K[\rho] = (E_{\text{ee}}[\rho] - J[\rho]) \quad (5.2.26)$$

Introducing the obvious normalization constraint on the electron density  $\int \rho(\mathbf{r})d\mathbf{r} = N$ , it should be easy to obtain

$$\begin{aligned} \frac{\delta}{\delta\rho(\mathbf{r})} \left[ E[\rho(\mathbf{r})] - \mu \int \rho(\mathbf{r})d\mathbf{r} \right] &= 0 \\ \Rightarrow \frac{\delta E[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})} &= \mu \end{aligned} \quad (5.2.27)$$

Equation (2.27) can now be written as the following by adding an effective potential term  $V_{\text{eff}}(\mathbf{r})$ :

$$\frac{\delta T_{\text{S}}[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})} + V_{\text{eff}}(\mathbf{r}) = \mu \quad (5.2.28)$$

where

$$V_{\text{eff}}(\mathbf{r}) = V_{\text{ext}}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + V_{\text{XC}}(\mathbf{r}) \quad (5.2.29)$$

and

$$V_{\text{XC}}(\mathbf{r}) = \frac{\delta E_{\text{XC}}[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})} \quad (5.2.30)$$

As we know, non-interacting electrons moving in the effective potential  $V_{\text{eff}}(\mathbf{r})$  would result in the same equation as Eq. (2.28). Thus, solving the one-electron Schrödinger equation is critical for us to determine the ground-state and its energy, which can be given as:

$$\left(-\frac{1}{2}\nabla_i^2 + V_{\text{eff}}(\mathbf{r}) - \varepsilon_i\right)\psi_i(\mathbf{r}) = 0 \quad (5.2.31)$$

Given that,

$$\rho(\mathbf{r}) = \sum_{i=1}^N |\psi_i(\mathbf{r})|^2 \quad (5.2.32)$$

Equations (2.29), (2.30), (2.31), and (2.32) can be solved self-consistently to obtain the ground-state energy for a given system, provided that the form of the term  $E_{\text{XC}}[\rho]$  is given empirically.

The advantage of DFT is quite obvious that it includes electron correlations with the computational cost of the HF method, but the disadvantage exists that the accuracy of the calculation is unpredictable because the exchange-correlation energy is empirical and a more complex form may or may not result in a better accuracy. It depends much on experience in giving a suitable form of the exchange-correlation energy regarding to the designated system and problem.

### 5.3 Basic Concepts of Equilibrium Statistical Physics

Before we move on to discuss how to implement MD simulations, it is critical for us to have some background knowledge about the basic concepts of equilibrium statistical physics, for the reason that most of the fundamental principles in MD simulations are related to them and that such understandings are crucially needed for correctly analyzing the results given by MD simulations.

It should be noted that, although computer simulation to a large extent provides researchers with a powerful tool to investigate properties of many-body systems, not all properties can be directly observed in computer simulations. Moreover, some of the quantities that can be directly measured in computer simulations in most cases cannot correspond to properties that experimentalists observe in real experiments. For instance, the direct results from MD simulations consist of instantaneous positions and velocities of different particles, which are microscopic information that cannot be directly compared with the properties measured in corresponding experiments. Normally, experiments measure average properties, which are averaged over either a large number of particles or a certain period of time, or both. Such a characteristic difference between MD simulations and experiments requires investigators to know what kind of averages to be computed. The following segment of this chapter provides a brief introduction on those averages that readers should have acquired before further delving into the detailed implementation of MD simulations.

### 5.3.1 Basic Concepts

Most of the computer simulations that the following segments are to talk about are based on that the classical statistical mechanics can be used in describing the detailed motions of atoms and molecules. This basis to a large extent simplifies most of the calculations, and is surprisingly appropriate when dealing with most cases in practice. Thus, the following section of this chapter is going to briefly talk about some of the basics of classical statistical physics.

Firstly, to understand the basics of thermodynamic and classical equilibrium statistical physics, it is vital to notice two basic hypotheses, *equiprobability* and *ergodicity*. *Equiprobability* states that, in a thermal system, there is an equal probability for the system to visit each microstate, whereas *ergodicity* states that all thermal states will be visited when the revolution time approaches infinity, namely, the ensemble average equals to the time average.

For further understanding, there are several concepts that need to be noted. For an  $N$  particles system with a total energy  $E$  that is confined in a volume  $V$ , we denote term  $\Omega(N, E, V)$  to indicate the number of all the microstates that it may visit, which also indicates its *ensemble*, namely, the set of all possible states under a given thermal condition. Now let us consider a system that is combined with two weakly interacting subsystems, which indicates that the total energy of the system follows the rule  $E = E_1 + E_2$ . Thus, for a given  $E_1$ , there is  $\Omega(E) = \Omega(E_1) \times \Omega(E_2)$ , and that

$$\ln \Omega(E, E - E_1) = \ln \Omega_1(E_1) + \ln \Omega_2(E - E_1) \quad (5.3.1)$$

Assuming that the two subsystems can transfer energy, then the most likely configuration of the distribution of the energy can be given with the equiprobability hypothesis, which is that the most likely value of  $E_1$  maximizes  $\ln \Omega(E, E - E_1)$ . Therefore,

$$\left( \frac{\partial \ln \Omega(E, E - E_1)}{\partial E_1} \right)_{N, V, E} = 0 \quad (5.3.2)$$

or

$$\left( \frac{\partial \ln \Omega_1(E_1)}{\partial E_1} \right)_{N_1, V_1} = \left( \frac{\partial \ln \Omega_2(E_2)}{\partial E_2} \right)_{N_2, V_2} \quad (5.3.3)$$

We can define

$$\beta(E, V, N) = \left( \frac{\partial \ln \Omega(E, V, N)}{\partial E} \right)_{N, V, E} \quad (5.3.4)$$

So Eq. (3.3) becomes



$$\beta(E_1, V_1, N_1) = \beta(E_2, V_2, N_2) \quad (5.3.5)$$

The maximum of  $\ln \Omega(N, E, V)$  also indicates that the system has found its equilibrium state, which coincides with the thermodynamic entropy  $S$  that also reaches the maximum when the system is at thermodynamic equilibrium. Thus, we can have the Boltzmann's entropic equation:

$$S = k_B \ln \Omega \quad (5.3.6)$$

where  $k_B$  is the Boltzmann constant. From Eq. (3.5), we can also see that  $\beta_1 = \beta_2$  when the two subsystems are at equilibrium, where it is easy to find the definition of the temperature  $T$ :

$$\frac{1}{T} = \left( \frac{\partial S}{\partial E} \right)_{N,V} \quad (5.3.7)$$

and then

$$\beta = \frac{1}{k_B T} \quad (5.3.8)$$

Another important concept that should be mentioned here is the *potential energy surface*, which indicates the set of potential energies for all configurations that a particular system may reach. These concepts will help us better understand not only computer simulations, but also the properties of many-body systems that we may be interested in.

### 5.3.2 Common Ensembles

We have introduced the concept of *ensemble* in the last segment, and to acquire knowledge about certain ensembles are of much significance in dealing with many-body systems and thus implementing computer simulations. We will briefly introduce several important ensembles, but to acquire further and detailed information, readers should check textbooks of statistical physics in order to develop a fundamental understanding of them.

A thermodynamic system with a constant particle number  $N$ , a constant energy  $E$ , and a constant volume  $V$  is defined as in the *microcanonical ensemble*. A system with a constant particle numbers  $N$ , a constant volume  $V$ , and a constant temperature  $T$  is defined as in the *canonical ensemble*. A system with a constant chemical potential  $\mu$ , a constant volume  $V$ , and a constant temperature  $T$  is defined to be in the *grandcanonical ensemble*, where the chemical potential follows:

$$\mu_i = \left( \frac{\partial U}{\partial N_i} \right)_{S, V, N_{j \neq i}} \quad (5.3.9)$$

which is defined by the phenomenological fundamental equation of thermodynamics, expressed in the form:

$$dU = TdS - PdV + \sum_i^n \mu_i dN_i \quad (5.3.10)$$

where label  $i$  differentiates different components of the system. Some other ensembles often used in computer simulations are the *isobaric-isothermal ensemble*, in which the particle number  $N$ , the pressure  $P$ , and the temperature  $T$  are constants, as well as the *isotension-isothermal ensemble*, which is adaptable to the shape of the simulation box.

The Boltzmann distribution of the canonical ensemble is

$$P_i = g_i \exp(-\beta E_i) / Z \quad (5.3.11)$$

where  $P_i$  stands for the probability of the microstate with energy  $E_i$ ,  $g_i$  stands for the degeneracy of energy state  $i$ , and the partition function  $Z$  is defined by

$$Z = \sum_i g_i \exp(-\beta E_i) \quad (5.3.12)$$

Another important feature that may need to be addressed here is the averages that different simulations essentially adopt during their implementations. For MC simulations, ensemble averages are implemented to generate an average of a certain observable  $A$

$$\langle A \rangle = \sum_i A_i g_i \exp(-\beta E_i) / Z \quad (5.3.13)$$

While for MD simulations, observables are often averaged over time:

$$\bar{A} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t A(t') dt' \approx \frac{1}{M} \sum_{i=1}^M A(t_i) \quad (5.3.14)$$

As we have discussed in the Sect. 3.1 about the hypothesis of ergodicity, the time average and ensemble average can both be applied in the computation of certain quantities. However, as the consequence of the fundamental difference of these simulations, there are certain distinctions between the two different simulations in practice, and understanding this is critical for researchers to understand the mechanism as well as the functionality and applicability of these simulations.

### 5.3.3 Common Thermodynamic Variables

There are certain thermodynamic variables and the relationships between them that are of much significance in dressing. Thus, in this segment, we are to introduce some of the important variables that researchers might encounter during MC or MD simulations. The detailed derivation and explanation of these variables are not the focus of this chapter. Interested readers may find them easily in any textbooks about statistical physics and thermodynamics.

In simulations, the *velocity*  $\mathbf{v}$ , *positon*  $\mathbf{r}$ , and applied *force*  $\mathbf{f}$  of the molecules are the most apparent observables. Many of the thermodynamic variables, unlike in real experiments, are deduced from them. First, we can give the definition of the *kinetic energy*:

$$E_k = \left\langle \sum_{i=1}^N \frac{1}{2} m_i \mathbf{v}_i^2 \right\rangle \quad (5.3.15)$$

The *temperature* can be deduced based on the Law of Equipartition of energy:

$$T = \frac{1}{dNk_B} \left\langle \sum_{i=1}^N m_i \mathbf{v}_i^2 \right\rangle \quad (5.3.16)$$

where  $d$  stands for the dimensionality of the given system. The *potential energy* can be derived as:

$$E_p = \left\langle \sum_{i=1}^N E_{pi} \right\rangle \quad (5.3.17)$$

The *pressure* can be calculated by

$$p = \frac{k_B T N}{V} - \frac{1}{dV} \left\langle \sum_{i < j} \mathbf{f}_{ij} \cdot \mathbf{r}_{ij} \right\rangle \quad (5.3.18)$$

Thus, the *enthalpy*, aka the total effective energy under the  $NPT$  ensemble is

$$H = E + pV \quad (5.3.19)$$

We have illustrated the *entropy* in Sect. 3.1 that

$$S = k_B \ln \Omega(N, V, E) \quad (5.3.20)$$

To illustrate how the system behaves under the  $NVT$  ensemble, we can define the *Helmholtz free energy*:

$$F = E - TS = -k_B T \ln Z \quad (5.3.21)$$

The *Gibbs free energy* can also be defined to demonstrate the system behavior for the *NPT* ensemble

$$G = F + pV = E - TS + pV \quad (5.3.22)$$

If the particle number  $N$  is not a constant, *chemical potential* can be used, as we have showed in Sect. 3.2:

$$\mu = \left( \frac{\partial G}{\partial N} \right)_{T,p} = \left( \frac{\partial F}{\partial N} \right)_{T,V} \quad (5.3.23)$$

From the definition of partition function  $Z$  in Eq. (3.12), it is also obvious that all the thermodynamic properties can be calculated from it as follows.

Energy:

$$U = -\frac{\partial}{\partial \beta} \ln Z \quad (5.3.24)$$

Free Energy:

$$F = -\frac{1}{\beta} \ln Z \quad (5.3.25)$$

Entropy:

$$S = -\frac{\partial F}{\partial T} \Big|_V = k_B \ln Z - k_B \beta \frac{\partial \ln Z}{\partial \beta} \quad (5.3.26)$$

Pressure:

$$P = -\frac{\partial F}{\partial V} \Big|_{T,N} = \frac{1}{\beta} \frac{\partial \ln Z}{\partial V} \Big|_{T,N} \quad (5.3.27)$$

Chemical potential:

$$\mu = \frac{\partial F}{\partial N} \Big|_{T,V} = -\frac{1}{\beta} \frac{\partial \ln Z}{\partial N} \Big|_{T,V} \quad (5.3.28)$$

Heat Capacity:

$$C_V = \frac{\partial U}{\partial T} \Big|_V = k_B \beta^2 \frac{\partial^2 \ln Z}{\partial \beta^2} = \frac{1}{T} \frac{\partial S}{\partial T} \Big|_V \quad (5.3.29)$$

## 5.4 Molecular Modeling

Before the implementation of a simulation, whether it is a Monte Carlo Simulation or a Molecular Dynamics Simulation, it is necessary to define how the simulation may be conducted. For all-atom simulations, the purpose of this is to treat a whole atom as a single particle, which reduces the computational complexity and simplify data analysis, leads to simulations with much larger temporal and spatial scales than first-principles calculations, and allows investigations on system behaviors under a finite temperature to be implemented. This requires the introduction of a *force field* describing the interactions among atoms, and the method to provide an empirical force field for a certain simulated system is called the molecular modeling method.

Normally, the molecular modeling method presets a functional form for the potential of two-body or many-body interactions with undetermined parameters, based on the characteristics of the interactions between atoms. The parameters can fitted either based on first-principles calculation data or experimental results. It should be noted that in an all-atom simulation, it is impossible to accurately depict all properties in a certain system when many degrees of freedom at the quantum level are reduced. Thus, it is of much importance for researchers to find out the set of parameters to determine, which coincides with the set of physical properties that researchers are mostly interested in during the simulation, and the accuracy for other properties can be loosened. The molecular model should be selected accordingly.

### 5.4.1 Reduced Unit

During the construction of certain molecular models and associated molecular simulations, it is useful to use the reduced unit instead of the SI-unit in the code to allow most variables taking numeric values not far from 1, and thus to minimize numeric truncation errors. The values can be recovered back to the SI-unit by multiplying a constant. The common conversion of the reduced unit is stated as follows.

The units of up to four basic physical variables are required to deduce all the units. If there are four units are given: length  $L$ , mass  $M$ , time  $t$ , and electric charge  $Q$ , we can calculate the units of all other physical variables as

Energy:

$$E = M * L^2 / t^2 \quad (5.3.30)$$

Temperature:

$$T = E / k_B \quad (5.3.31)$$

Pressure:

$$P = E/L^3 \quad (5.3.32)$$

Mass Density:

$$\rho = M/L^3 \quad (5.3.33)$$

Molecular Number Density:

$$n = 1/L^3 \quad (5.3.34)$$

Dielectric Constant:

$$\varepsilon = \frac{N_A \cdot Q^2}{L \cdot E} \quad (5.3.35)$$

### 5.4.2 Lennard-Jones Potential

The most widely used potential to describe the van der Waals (VDW) interaction is the Lennard-Jones (LJ) potential. The most popular one is called **12-6 LJ**:

$$V(r) = 4\varepsilon \left( \frac{\sigma^{12}}{r^{12}} - \frac{\sigma^6}{r^6} \right) \quad (5.3.36)$$

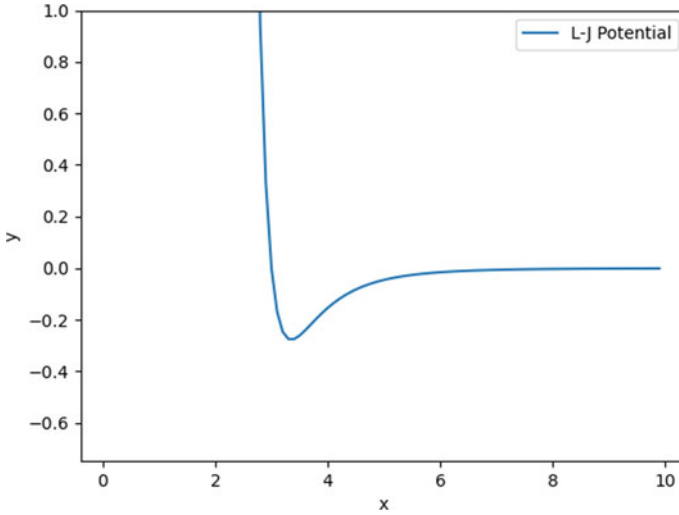
The corresponding force is

$$\mathbf{F}(r) = -\nabla V(r) = 24 \frac{\varepsilon}{r} \left( 2 \frac{\sigma^{12}}{r^{12}} - \frac{\sigma^6}{r^6} \right) \hat{\mathbf{r}} \quad (5.3.37)$$

For an inertial gas, the interaction between gas atoms can be well described by the LJ potential, as shown in Fig. 5.1. The LJ potential are also widely used to describe the non-bonded VDW interactions between atoms in chemical or biological systems.

The accumulation of the LJ potential for the part larger than a *cutoff distance*  $r_c$  can be given as a constant. To demonstrate this, we can have

$$\int_{r_c}^{\infty} V(r) 4\pi r^2 dr \quad (5.3.38)$$



**Fig. 5.1** A typical form of the Lennard-Jones potential

An interaction is *short-range* if the integral Eq. (3.38) converges, and *long-range* otherwise. The LJ potential is obviously a short-range interaction, and we can regard the LJ force as zero when the distance  $r > r_c$ .

### 5.4.3 Force Fields for Metals

We usually have to describe metals by many-body force fields, since the valence electrons can move globally in due systems. A general model that can describe both pure metal and alloy is the Embedded Atom Model (EAM), the potential of which is given as

$$V_i = F_\alpha \left( \sum_{i \neq j} \rho_\alpha(r_{ij}) \right) + \frac{1}{2} \sum_{i \neq j} \phi_{\alpha\beta}(r_{ij}) \quad (5.3.39)$$

In Eq. (3.39), the distance  $r_{ij}$  refers to the distance between atom  $i$  and atom  $j$ . The term  $\phi_{\alpha\beta}$  refers to the two-body potential between two metal atoms of type  $\alpha$  and type  $\beta$ , respectively. The term  $\rho_\alpha$  represents the amount of electron density that atom  $j$  of type  $\alpha$  generates at the position where atom  $i$  locates, whereas function  $F_\alpha$  refers to an embedded function that represents the energy needed to embed atom  $i$  of type  $\alpha$  into the electron cloud.

There are various other force fields for metals whose functional forms are usually a many-body functional term of electron density plus a two-body term.

### 5.4.4 Chemical and Biomolecular Force Fields

If we do not consider the effects brought by many-body molecular polarization, a general approach to depict chemical and biomolecular systems is to divide the interactions into two parts. One is the *bonded interactions* and the other is *non-bonded interactions*. For bonded interactions, there are valence bonds, valence angles, dihedral angles, and improper dihedral angles that we should take into consideration. For non-bonded interactions, the VDW and electrostatic interactions are ones that we must depict in the potential. These are schematically illustrated in Fig. 5.2.

To demonstrate how the depiction of these interactions is made, we can refer to the AMBER all-atom force field as follows, where  $k_r$ ,  $r_0$ ,  $k_\theta$ ,  $\theta_0$ ,  $V_n$ ,  $n$ ,  $\gamma$ ,  $\epsilon$ , and  $\sigma$  are parameters that are yet to be determined either through first-principles calculations or through experiments.

#### 1. Bonded Interactions:

- (a) Valence Bond: normally a covalent bond between two atoms depicted by a harmonic oscillator:

$$V_{\text{bond}} = \frac{1}{2}k_r(r - r_0)^2 \quad (5.3.40)$$

- (b) Valence Angle: the angle between three adjacent atoms depicted by a harmonic oscillator:

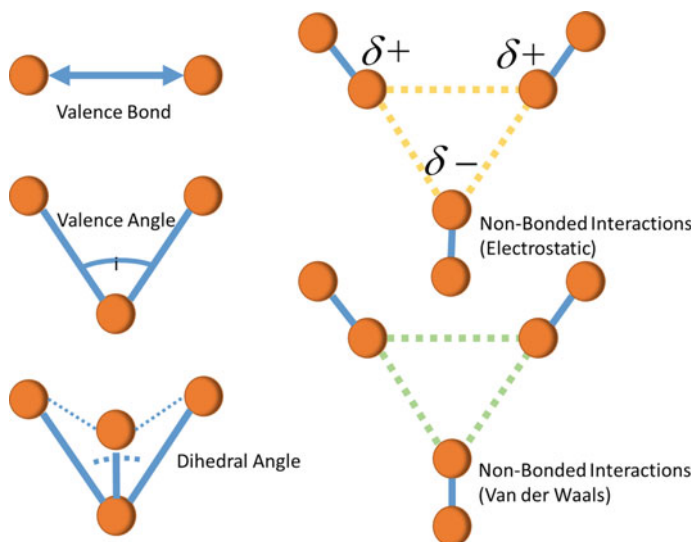


Fig. 5.2 Bonded and non-bonded interactions



$$V_{\text{angel}} = \frac{1}{2}k_{\theta}(\theta - \theta_0)^2 \quad (5.3.41)$$

- (c) Dihedral Angle: the dihedral angle between four sequential atoms, depicted in the form of:

$$V_{\text{dihedral}} = V_n \cos(n\phi - \gamma) \quad (5.3.42)$$

- (d) Improper Dihedral Angle: the angle that constrains four atoms in one plane, which could also be depicted in the form of equation Eq. (3.42).

## 2. Non-bonded Interactions:

- (a) VDW interactions: as we have mentioned in Sect. 4.2, the VDW interactions can be depicted by the LJ-potential:

$$V_{\text{LJ}} = 4\epsilon \left( \frac{\sigma^{12}}{r^{12}} - \frac{\sigma^6}{r^6} \right) \quad (5.3.43)$$

- (b) Electrostatic interactions:

$$V_{\text{EL}} = \frac{Kq_1q_2}{r} = \frac{1}{4\pi\epsilon_0} \frac{q_1q_2}{r} \quad (5.3.44)$$

Therefore, the interacting potential of a chemical or biomolecular system described by the AMBER force field can be given as:

$$\begin{aligned} V = & \sum_{\text{bonds}} k_r(r - r_0)^2 + \sum_{\text{angles}} k_{\theta}(\theta - \theta_0)^2 + \sum_{\text{torsions}} \sum_n \frac{1}{2}V_n[1 + \cos(n\phi - \phi_0)] \\ & + \sum_{j=1}^{N-1} \sum_{i=j+1}^N \epsilon_{ij} \left[ \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^6 \right] + \sum_{j=1}^{N-1} \sum_{i=j+1}^N \frac{1}{4\pi\epsilon_0} \frac{q_iq_j}{r_{ij}} \end{aligned} \quad (5.3.45)$$

There are also other popular general-purpose all-atom force fields with different appliances, such as CHARMM, OPLS, COMPASS, DRUDE, etc., whose mathematical expressions are slightly different from each other because they emphasize different aspects of chemical or biomolecular systems.

### 5.4.5 Coarse-Graining Methodology

Analogous to emerging atomistic force fields from the quantum level, the coarse-graining (CG) methodology is intended to enhance the simulation temporal and spatial scales by reducing the degrees of freedom at the atomistic level. Difficulties

come from the fact that, unlike reducing the degrees of freedom at the quantum level, the adiabatic assumption usually does not work well at the atomistic level. Moreover, similar to atomistic force fields, different CG methods emphasize different physical properties and loosen the accuracy for other properties. A CG method normally requires either fitting from experimental and/or atomistic simulation data or converting mathematically from the atomistic force field. Two typical CG methods of the latter are the Multiscale Coarse-Graining (MS-CG) method and the Effective Force Coarse-Graining (EF-CG) method, both present researchers with tools that reduce degrees of freedom and accelerate dynamics while containing necessary physical properties in chemical and biomolecular systems. The MARTINI CG force field is a widely used general-purpose one developed for biomolecular systems. Detailed explanations on their mechanism and applicability can be found in Refs. [4–6].

## 5.5 Monte Carlo (MC) Simulation

### 5.5.1 Purpose

In the previous sections, we have brought forward some basic concepts of thermodynamics and classical statistical mechanics. To introduce the MC method, we can start from the classical expression of the partition function  $Z$ , which has been introduced in Sect. 3.2. For a system with  $N$  identical atoms in the canonical ensemble, the partition function becomes:

$$Z = \frac{1}{(h^{dN} N!)} \int d\mathbf{p}^N d\mathbf{r}^N \exp[-\beta E(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3 \dots \mathbf{r}_N, \mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3 \dots \mathbf{p}_N)] \quad (5.5.1)$$

where  $\mathbf{r}_i$  and  $\mathbf{p}_i$  stand for the coordinates and momenta of the particle labeled  $i$ . The corresponding ensemble average of a certain observable  $A$  thus becomes

$$\langle A \rangle = \frac{\int \exp(-\beta E(\mathbf{r}^N, \mathbf{p}^N)) A(\mathbf{r}^N, \mathbf{p}^N) d\mathbf{r}^N d\mathbf{p}^N}{\int \exp(-\beta E(\mathbf{r}^N, \mathbf{p}^N)) d\mathbf{r}^N d\mathbf{p}^N} \quad (5.5.2)$$

It is obvious that the observable  $A$  is written as the function of coordinates and momenta. From the kinetic energy Eq. (3.15), we can see that the integration over the momenta can be carried out analytically. However, the computation of the average of function  $A(\mathbf{r}^N)$  in the position space is difficult, and the multidimensional integral over the coordinates can only be analytically calculated in the simplest cases. Thus, it requires the development of certain numerical techniques to do the integration in the position space. One of the techniques that is well developed and widely used is the MC method, which was first brought up by Metropolis et al. in 1953 [7]. In this section, we will majorly focus on this method and its implementation.

### 5.5.2 Importance Sampling

Before moving on to discuss importance sampling, it is useful to have a look at the simplest sampling, random sampling, which will allow a clear understanding of the MC method. If we want to numerically solve a one-dimensional integral:

$$I = \int_a^b f(x) dx \quad (5.5.3)$$

It is clear that we can rewrite (5.3) into

$$I = (b - a) \langle f(x) \rangle \quad (5.5.4)$$

In random sampling, the average  $\langle f(x) \rangle$  is determined by evaluating  $f(x)$  at a very large number  $L$  of variable  $x$ , which are randomly and uniformly distributed in the interval  $[a, b]$ . But in cases where we want to solve equations like Eq. (5.2), it is very inefficient to adopt random sampling, and the specific feature of the Boltzmann distribution allows high efficiency of importance sampling. The basic concept of importance sampling is to randomly sample in the configurational space in such a way that the configurations contributed the most to the ensemble average are sampled and those have negligible contributions are not sampled.

A simple demonstration can be used to show how this might be achieved. We can compute the one-dimensional integral mentioned earlier, but with the sampling points distributed nonuniformly over the interval  $[0, 1]$  according to a nonnegative probability density  $p(x)$ . Thus, we can rewrite Eq. (5.3) to be

$$I = \int_0^1 dx \left( p(x) \frac{f(x)}{p(x)} \right) \quad (5.5.5)$$

Assuming that the function  $p(x)$  is the derivative of a nonnegative, nondecreasing function  $u(x)$ , and that  $p(x)$  is normalized. We have

$$I = \int_0^1 du \frac{f([x(u)])}{p([x(u)])} \quad (5.5.6)$$

Then by generating  $L$  random values of  $u$  uniformly distributed within the interval  $[0, 1]$ , similar to what we have done in Eq. (5.4), we can have the estimation for  $I$  as

$$I \approx \frac{1}{L} \sum_{i=1}^L \frac{f[x(u_i)]}{p[x(u_i)]} \quad (5.5.7)$$

Thus, we can conduct the importance sampling to a certain integral. This is of much significance in the MC method, which in its essence performs the importance sampling for the Boltzmann distribution. Instead of choosing configurations randomly then weighting them with  $\exp(-E/kT)$ , configurations are chosen with a probability  $\exp(-E/kT)$  and then weighted evenly. The most popular implementation of the MC method is the Metropolis algorithm, which will be illustrated in the next segment.

### 5.5.3 Metropolis Algorithm

In most cases, we are not interested in the absolute value of the configuration part of the partition function, but the averages of some observables in the form of

$$\langle A \rangle = \frac{\int \exp(-\beta U(\mathbf{r}^N)) A(\mathbf{r}^N) d\mathbf{r}^N}{\int \exp(-\beta U(\mathbf{r}^N)) d\mathbf{r}^N} \quad (5.5.8)$$

That is to say, the main focus is to figure out the ratio of two integrals. It is possible to calculate such as ratio through the Metropolis algorithm [7]. The configuration part of the partition function is

$$Q \equiv \int \exp(-\beta U(\mathbf{r}^N)) d\mathbf{r}^N \quad (5.5.9)$$

We can denote the probability density by defining

$$N_Q(\mathbf{r}^N) \equiv \frac{\exp(-\beta U(\mathbf{r}^N))}{Q} \quad (5.5.10)$$

To do the sampling, we can assume that we have  $L$  total sampling points generated randomly in the configurational space according to the probability density  $N_Q(\mathbf{r}^N)$ , and among them there are  $n_i$  points generated around a point  $\mathbf{r}^N$ . Thus we have

$$\langle A \rangle \approx \frac{1}{L} \sum_{i=1}^L n_i A(\mathbf{r}_i^N) \quad (5.5.11)$$

The problem now becomes how to generate such sampling points in the configurational space with the probability distribution of the Boltzmann distribution. To do this, we can firstly prepare the system in the configuration  $\{\mathbf{r}^N\}$ , denoted by  $o$ , having a Boltzmann factor  $\exp[-\beta U(o)]$ . Then we can add some small displacement to the original configuration and generate the trail configuration  $\{\mathbf{r}^N\}$ , denoted by  $n$ , following  $\exp[-\beta U(n)]$ . Whether to accept this trail configuration or not is the key where Metropolis scheme can be applied, which will be demonstrated as follows.

It should be noted from the detailed balance condition that the transition probability from configuration  $n$  to  $o$ , or  $\pi(n \rightarrow o)$ , follows an equation that

$$N_Q(o)\pi(o \rightarrow n) = N_Q(n)\pi(n \rightarrow o) \quad (5.5.12)$$

In practice, we can construct  $\pi(o \rightarrow n)$  with transition matrix  $\alpha(o \rightarrow n)$ , which is the underlying matrix for the corresponding Markov chain, and accepting probability  $acc(o \rightarrow n)$  so that:

$$\pi(o \rightarrow n) = \alpha(o \rightarrow n) \times acc(o \rightarrow n) \quad (5.5.13)$$

If matrix  $\alpha$  is symmetric, we can know that

$$\frac{acc(o \rightarrow n)}{acc(n \rightarrow o)} = \frac{N_Q(n)}{N_Q(o)} = \exp\{-\beta[U(n) - U(o)]\} \quad (5.5.14)$$

We can choose  $acc(o \rightarrow n)$  to fulfill the condition Eq. (5.14) by assigning

$$acc(o \rightarrow n) = \begin{cases} N_Q(n)/N_Q(o) & \text{if } N_Q(n) < N_Q(o) \\ 1 & \text{if } N_Q(n) \geq N_Q(o) \end{cases} \quad (5.5.15)$$

So that the transition probability from state  $o$  to state  $n$  is given by

$$\pi(o \rightarrow n) = \begin{cases} \alpha(o \rightarrow n) & \text{if } N_Q(n) \geq N_Q(o) \\ \alpha(o \rightarrow n)[N(n)/N(o)] & \text{if } N_Q(n) < N_Q(o) \end{cases} \quad (5.5.16)$$

$$\pi(o \rightarrow o) = 1 - \sum_{o \neq n} \pi(o \rightarrow n)$$

To decide whether to accept such a trial move or not, we can generate a random number  $X$  from a uniform distribution between  $[0, 1]$ . We can then accept the trial move if  $X$  is less than or equals to  $acc(o \rightarrow n)$  and reject it otherwise. This gives the eventual sequence of how we can perform an MC simulation.

The procedure of an MC simulation with the Metropolis algorithm can then be given as follows. First, generate a new “trial” configuration by making a perturbation to the present configuration. Then, accept the new configuration based on the ratio of the probabilities for the new and old configurations, according to the Metropolis algorithm. If the trial is rejected, the present configuration is taken as the next one in the Markov chain. The above steps are repeated many times, and instantaneous data and configurations are sampled for data analysis.

From what we have discussed we can have a snapshot of the characteristics of MC simulation. By default, the MC simulation simulates a system under the  $NVT$  ensemble. The potential energy is calculated without force, which saves some computational complexity and time. It is mostly used for simulating equilibrium systems,

for the reason that it is based on the importance sampling of the equilibrium Boltzmann distribution. The simulation step can be arbitrarily defined for the integration in Eq. (5.8). The MD simulations are better to be used if we want to study non-equilibrium systems, or to reproduce microscopic dynamics of particles in the systems.

The default MC simulation implements the *NVT* ensemble. However, in practice, it is necessary for us to implement different ensembles, for which a series of techniques are often applied. As we have previously suggested, the key of the algorithm is to use a Monte Carlo procedure that introduces a random walk in the regions of phase space that contribute to the ensemble average. The acceptance rules are determined to attain the needed probability distribution, where the detailed balance condition is fundamental. Thus, to sample different distributions according to different ensembles, we can use the following procedure:

1. Determine the distribution that needed to be sampled.
2. Impose the detailed balance condition:

$$M(o \rightarrow n) = M(n \rightarrow o) \quad (5.5.17)$$

where  $M(o \rightarrow n)$  refers to the flow from configuration  $o$  to  $n$ , which is the product of the probability of the emergence of configuration  $n$  from configuration  $o$ , which is  $\alpha(o \rightarrow n)$ , and the acceptance probability  $acc(o \rightarrow n)$ :

$$M(o \rightarrow n) = \alpha(o \rightarrow n) \times acc(o \rightarrow n) \quad (5.5.18)$$

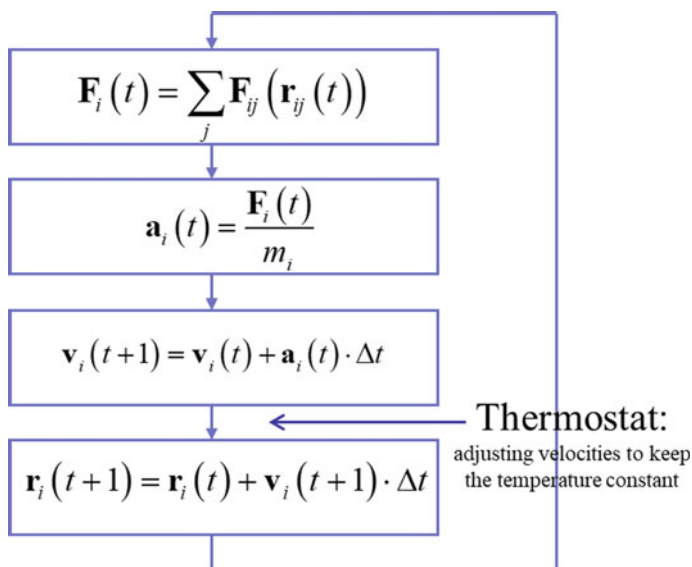
3. Determine the probability for a particular configuration
4. Come up with the condition that acceptance rules should meet.

This procedure is highly general, the details of which may vary in accordance with the ensemble that researchers are interested in. The readers should be clear that one should be extremely cautious when treating different ensembles with the MC method, since errors may easily be introduced. The choice of ensembles for Monte Carlo simulations is very broad. Isobaric-isothermal, constant-stress-isothermal, grand canonical, and microcanonical ensembles are all suitable. For further information, one may refer to Refs. [8–11].

## 5.6 Molecular Dynamics (MD) Simulation

### 5.6.1 Idea of Molecular Dynamics

The substantial idea of MD is to solve classical many-body Newton's equations of motion numerically. The typical procedure of a MD simulation is presented in Fig. 5.3. It should be noted that how to obtain the empirical force field  $\{\mathbf{F}_{ij}\}$  is vital



**Fig. 5.3** Procedure of a typical MD simulation

in conducting the simulation in certain systems. Normally, the force field  $\{\mathbf{F}_{ij}\}$  is determined by fitting experimental and/or first-principles calculation data, and the determination highly influences the quality of the simulation. In practice, with this method, researchers are allowed to simulate millions of atoms for tens of nanoseconds with parallel computing. Compared to the simulation conducted by the first-principles calculation methods, it allows simulations with a larger time span to happen.

There are certain main characteristics that MD simulation have, allowing it to be applicable to many research fields. First, it allows researchers to simulate real experiments, which enables detailed guidance to experimentalists and comparison between theory and experiments. Second, the averages of certain observables are time averages other than ensemble averages, and the time steps are limited compared to the MC simulations. Thus, it is applicable to simulate non-equilibrium cases and show the dynamics of some properties of the given system. Finally, it calculates both potential energy and interactive forces in between the molecules. There are several MD software packages available nowadays, such as AMBER, GROMACS, LAMMPS, CHARMM, and NAMD, etc. To better conduct MD simulation, it is essential to understand the mechanism of such simulation. The detailed simulation, sampling, as well as the basic data analysis of it will be showed in the following part of this chapter. Of course, it is impossible for us to discuss this subject in merely a section. To further investigate into this, readers can refer to Refs. [8, 12, 13].

## 5.6.2 MD Simulation and Sampling

As we have discussed earlier, the MD method simulates many-body systems through solving many-body Newton's Equations of Motion by calculating the resultant force on each particle. The basic process of the MD simulation is illustrated in Fig. 5.3. Some details of the simulation algorithms and related concepts will be illustrated in the following parts.

### 5.6.2.1 Basic Concepts

Before digging into details of the simulation process, it is beneficial to clarify some basic concepts that are normally introduced in MD simulations.

It is understandable that the simulation should be conducted in a limited space within a limited time scale. Thus, the simulation space should be given first. The *continuity of simulation space* can be either discrete, for instance, the space for the Ising Model, or continuous. The *Boundary Condition* for the space can be either free, rigid, or periodic. Among them, the most widely used and thus most important boundary condition is the *periodic boundary condition* (PBC), under which particles in the simulated box can interact with an infinite number of their images, allowing us to study physical properties of macroscopic systems with  $\sim 10^{23}$  particles by simulating only  $\sim 10^3$  to  $\sim 10^6$  particles. The total energy of a system simulated with the PBC becomes:

$$E_{\text{tot}} = \frac{1}{2} \sum'_{i,j,\mathbf{n}} E(|\mathbf{r}_{ij} + \mathbf{n}L|) \quad (5.6.1)$$

where  $L$  represents the length of the simulation box,  $\mathbf{n}$  is a three-dimensional integer vector, and the summation  $\sum'$  excludes the addition of the term  $\mathbf{n} = 0$  when  $i = j$ . During the simulation with the PBC, particles moving out of the simulation box will be placed back into the simulation box from the other side of the simulation box, conceptually mimicking that, when one particle moves out of the simulation box, there is an equal probability that another particle moves into it in the symmetric way. The simulation box length  $L$  must be set longer than twice of the *cutoff distance* for the longest short-range force in the real space to avoid the interaction of a particle with two of its imaging particles simultaneously, which means that

$$L > 2r_c \quad (5.6.2)$$

The long-range interaction must be calculated with the *Ewald summation* method that will be introduced later.

There are three major approaches to numerically treat the potential with a *cutoff distance*. The first is simply setting the potential beyond the cutoff distance to be zero:



$$V(r) = \begin{cases} V^{\text{real}}(r) & r \leq r_c \\ 0 & r > r_c \end{cases} \quad (5.6.3)$$

where  $V^{\text{real}}$  is the original mathematical expression of the potential. This treatment is simple but has two problems. The first is that the absolute value of the potential has a jump and correspondingly the force diverges at the cutoff distance. The second is that there is an artificial contribution to the system pressure due to the discontinuity of the potential. To avoid the above problem, another approach constructs the potential in the following way:

$$V(r) = \begin{cases} V^{\text{real}}(r) - V^{\text{real}}(r_c) & r \leq r_c \\ 0 & r > r_c \end{cases} \quad (5.6.4)$$

which is the common method in dealing with the cutoff potential. The third way treats the potential with a cutoff distance by introducing a smooth switching function that truncates the potential energy smoothly at the cutoff distance, tapering the interaction potential over a predefined range of distances. The potential approaches zero smoothly between the first and last cutoff but remains the usual value before the first cutoff.

Another important concept that should be taken into consideration in defining the simulation space is the *characteristic length*, which is the spatial correlation length of a certain physical property. In principle, the simulation box size should be larger than the characteristic lengths of important properties. In practice, this condition is not always able to be satisfied, then the finite-size effect caused by the insufficient size of the simulation box can be studied by varying the simulation size and observing the limiting behavior towards infinity. The characteristic length of a physical variable  $A$  can be determined by calculating the spatial correlated function:

$$c(\mathbf{r}) = \langle \delta A(\mathbf{r}_0 + \mathbf{r}) \delta A(\mathbf{r}_0) \rangle \quad (5.6.5)$$

where  $\delta A(\mathbf{r}) = A(\mathbf{r}) - \langle A(\mathbf{r}) \rangle$ .

For sampling, theoretically the importance sampling within a limited simulation duration ensures that the sampled data contribute most to the ensemble average. Because either MC or MD simulation procedure already generates data with the weight of the Boltzmann distribution, a uniform time interval should be employed to sample simulation data. The initial configuration should be set to be as close as possible to the equilibrium state. Normally a pre-simulation procedure not be sampled is required to equilibrate the system from the usually non-equilibrated initial configuration. Another significant factor that we should consider is the correlation of sampled data. A closer sampling results in a larger correlation, and the standard deviation of the sampled data becomes smaller when the data are more correlated. Thus when the fluctuations of thermodynamic properties are concerned, the sampling interval should in principle be set as large as possible under the condition that enough number of data are sampled.

### 5.6.2.2 Numerical Integration of Newton's Equations of Motion

Regarding calculating the integral of Newton's Equations of Motion numerically, the key consideration is accuracy and long-time stability, while computational cost is negligible. Below we introduce several commonly used numerical integration methods.

#### Verlet Algorithm

The *Verlet algorithm* is the basis for a set of algorithms. To implement this, we should first do the Taylor expansion to the Newton's Equations of Motion, so that

$$\begin{aligned}
 r(t + \Delta t) &= r(t) + v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 + \frac{\Delta t^3}{3!}\ddot{r} + O(\Delta t^4) \\
 r(t - \Delta t) &= r(t) - v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 - \frac{\Delta t^3}{3!}\ddot{r} + O(\Delta t^4) \\
 \Rightarrow r(t + \Delta t) + r(t - \Delta t) &= 2r(t) + \frac{f(t)}{m}\Delta t^2 + O(\Delta t^4) \\
 \Rightarrow r(t + \Delta t) &\approx 2r(t) - r(t - \Delta t) + a\Delta t^2
 \end{aligned} \tag{5.6.6}$$

The velocity can then be estimated by

$$v(t) = \frac{r(t + \Delta t) - r(t - \Delta t)}{2\Delta t} + O(\Delta t^2) \tag{5.6.7}$$

#### Velocity Verlet Algorithm

Another algorithm is called the *velocity Verlet algorithm*, which is of better accuracy and stability compared to the Verlet algorithm, but cannot be used along with the Nosé-Hoover thermostat. To do so, we can have

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 = r(t) + v(t)\Delta t + \frac{a}{2}\Delta t^2 \tag{5.6.8}$$

Then the velocity can be determined by

$$v(t + \Delta t) = v(t) + \frac{f(t + \Delta t) - f(t)}{2m}\Delta t = v(t) + \frac{1}{2}[a(t + \Delta t) + a(t)]\Delta t \tag{5.6.9}$$

### Leap Frog Algorithm

The *leap frog algorithm* calculates velocities in between time steps, which allows it to be suitable for cooperating with some thermostat algorithms.

If we define:

$$\begin{aligned} v\left(t - \frac{\Delta t}{2}\right) &\equiv \frac{r(t) - r(t - \Delta t)}{\Delta t} \\ v\left(t + \frac{\Delta t}{2}\right) &\equiv \frac{r(t + \Delta t) - r(t)}{\Delta t} \end{aligned} \quad (5.6.10)$$

Then we have:

$$\begin{aligned} r(t + \Delta t) &= r(t) + v\left(t + \frac{\Delta t}{2}\right) \cdot \Delta t \\ v\left(t + \frac{\Delta t}{2}\right) &= v\left(t - \frac{\Delta t}{2}\right) + a \cdot \Delta t \end{aligned} \quad (5.6.11)$$

### Prediction-Correction Algorithm

The *prediction-correction algorithm* performs the integration by using the mathematical prediction-correlation method. It is asymmetric in time, namely, the inverse of the time in doing integration does not bring the system back to the beginning point.

Let us start with the Taylor expansion:

$$r(t + \Delta t) = r(t) + \frac{\partial r}{\partial t} \Delta t + \frac{\partial^2 r}{\partial t^2} \frac{\Delta t^2}{2!} + \frac{\partial^3 r}{\partial t^3} \frac{\Delta t^3}{3!} + \dots \quad (5.6.12)$$

We can define:

$$\begin{aligned} x_0(t) &\equiv r(t) \\ x_1(t) &\equiv \frac{\partial r}{\partial t} \Delta t \\ x_2(t) &\equiv \frac{\partial^2 r}{\partial t^2} \frac{\Delta t^2}{2!} \\ x_3(t) &\equiv \frac{\partial^3 r}{\partial t^3} \frac{\Delta t^3}{3!} \end{aligned} \quad (5.6.13)$$

Thus, we have the prediction that:

$$\begin{cases} x_0^{\text{predicted}}(t + \Delta t) = x_0(t) + x_1(t) + x_2(t) + x_3(t) \\ x_1^{\text{predicted}}(t + \Delta t) = x_1(t) + 2x_2(t) + 3x_3(t) \\ x_2^{\text{predicted}}(t + \Delta t) = x_2(t) + 3x_3(t) \\ x_3^{\text{predicted}}(t + \Delta t) = x_3(t) \end{cases} \quad (5.6.14)$$

Because we can calculate the actual acceleration  $x_2^{\text{corrected}}$  with Newton's second law, we can correct the second term:

$$\Delta x_2 = x_2^{\text{corrected}} - x_2^{\text{predicted}} \quad (5.6.15)$$

Then we can give corrections to the other terms accordingly:

$$x_n^{\text{corrected}} = x_n^{\text{predicted}} + c_n \Delta x_2 \quad (5.6.16)$$

For the algorithm of a certain order, the constants  $c_n$  should be so chosen that the accuracy and the long-time stability of the calculation are well balanced. Theoretically, the predictions and corrections should be calculated recursively to make the results self-consistent, but in practice, because each step of recursion requires time-consuming calculations of the force, a better way to improve the accuracy is reducing the step length and performing the calculation only once at each step.

### 5.6.2.3 Thermostat (or Heat Bath)

The default MD simulation realizes systems with the microcanonical ensemble, or *NVE* ensemble, as we have demonstrated in Sect. 3.2. Because most of the time researchers are more interested in simulating with a constant-temperature ensemble, such as the *NVT* or *NPT* ensemble, it is necessary to introduce the *thermostat* (*heat bath*) algorithm after the integration of the Newton's equations of motion to adjust the velocities of particles, so that the system temperature can be kept statistically a constant.

#### Isokinetics Thermostat

The *isokinetics thermostat* resets the system temperature directly to the designated one at the interval of one or multiple steps by brutally rescaling all velocities:

$$\frac{3}{2} N k_B T = \frac{1}{2} \sum_i m_i v_i^2 \Rightarrow v_i^{\text{scale}} = \lambda v_i, \lambda = \sqrt{\frac{T}{T_0}} \quad (5.6.17)$$

The problem of this resides in the fact that, although the temperature is fixed to a constant, the fluctuation of the temperature disappears, meaning that it does not

reflect the real physics of the canonical ensemble in the momentum space. However, it can normally provide correct statistics in the position space.

### Berendsen Thermostat

Instead of rescaling the system temperature directly to the designated one, the *Berendsen thermostat* [14] allows the system to relax to the designated temperature with a characteristic relaxation time  $\tau_T$  in multiple steps. The definition of  $\lambda$  in Eq. (6.17) is replaced by

$$\lambda = \left[ 1 + \frac{\Delta t}{\tau_T} \left( \frac{T}{T_0} - 1 \right) \right]^{\frac{1}{2}} \quad (5.6.18)$$

which is equivalent to:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau_T} \quad (5.6.19)$$

This thermostat introduces a certain degree of fluctuation but the physics in the momentum space is still incorrect.

### Andersen Thermostat

The *Andersen thermostat* proposed by Andersen [15] can achieve the proper simulation of the momentum space in the canonical ensemble by introducing stochastic collisions of thermostat to particles.

If two successive collisions are uncorrelated, it is obvious that the times the collisions happen during time interval  $t$  with a collision frequency  $\lambda$  follow the Poisson distribution:

$$P(t; \lambda) = \lambda \exp(-\lambda t) \quad (5.6.20)$$

Therefore, a typical procedure to apply the Andersen thermostat in implementing MD simulations has the following steps:

1. calculate the regular MD integration of Newton's Equations of Motion with a time interval of  $\Delta t$ ;
2. a set of particles are chosen randomly with the probability of choosing each particle to be  $\lambda \Delta t$ ;
3. for each chosen particle, a velocity is generated randomly according to the Maxwell-Boltzmann distribution under temperature  $T$  and assigned to the particle to replace its current velocity.

It can be proven that the Andersen thermostat indeed simulates the momentum space of the canonical ensemble correctly. However, the angular momentum of the system is not conserved during the replacement of particle velocities, causing the whole system to rotate randomly in accordance with time, which leads to some difficulties for conducting data analysis.

### Nosé-Hoover Thermostat

The basic idea of the Nosé-Hoover thermostat [15–19] is to extend the Lagrangian of the system by adding additional, fictitious position and velocity as extra degrees of freedom, making the extended system to be in the microcanonical ensemble while the actual system is in the canonical ensemble.

As we know, the Lagrangian of the system is

$$L = E_k - E_p \quad (5.6.21)$$

and the Hamiltonian of the system:

$$H = E_k + E_p \quad (5.6.22)$$

The expanded Lagrangian is defined as

$$L_{\text{Nose}} = \sum_{i=1}^N \frac{m_i}{2} s^2 \dot{\vec{r}}_i^2 - E_p(\vec{r}^N) + \frac{Q}{2} \dot{s}^2 - \frac{l}{\beta} \ln s \quad (5.6.23)$$

where  $s$  is the additional position,  $l$  is an unidentified parameter, and  $Q$  is the “effective mass” for position  $s$ . The momentum can thus be defined as

$$\begin{aligned} \vec{p}_i &\equiv \frac{\partial L}{\partial \dot{\vec{r}}_i} = m_i s^2 \dot{\vec{r}}_i = m_i s^2 \frac{d\vec{r}_i}{dt} \\ \vec{p}_s &\equiv \frac{\partial L}{\partial \dot{s}} = Q \dot{s} = Q \frac{ds}{dt} \end{aligned} \quad (5.6.24)$$

Accordingly, the expanded Hamiltonian becomes

$$H_{\text{Nose}} = \sum_{i=1}^N \frac{\vec{p}_i^2}{2m_i s^2} + E_p(\vec{r}^N) + \frac{p_s^2}{2Q} + \frac{l}{\beta} \ln s \quad (5.6.25)$$

With this method, we can expand the system with  $N$  particles to a system with  $6N + 2$  degrees of freedom. We then prove below that the expanded system can realize the simulation of the actual system in the canonical ensemble correctly.

If we define  $\vec{p}'_i \equiv \frac{\vec{p}_i}{s}$ , the partition function can be given as

$$\begin{aligned}
Z_{\text{Nose}} &= \frac{1}{N!} \int dp_s ds d\vec{p}^N d\vec{r}^N \delta(H_{\text{Nose}} - E) \\
&= \frac{1}{N!} \int dp_s ds d\vec{p}^N d\vec{r}^N \delta\left(\sum_{i=1}^N \frac{\vec{p}_i^2}{2m_i} + E_p(\vec{r}^N) + \frac{p_s^2}{2Q} + \frac{l}{\beta} \ln s - E\right) \\
&= \frac{1}{N!} \int dp_s ds d\vec{p}^N d\vec{r}^N \delta(h)
\end{aligned} \tag{5.6.26}$$

The Hamiltonian of the actual system is

$$H(\vec{p}', \vec{r}) = \sum_{i=1}^N \frac{\vec{p}_i'^2}{2m_i} + E_p(\vec{r}^N) \tag{5.6.27}$$

Thus, we have

$$h(s) = H(\vec{p}', \vec{r}) + \frac{p_s^2}{2Q} + \frac{l}{\beta} \ln s - E \tag{5.6.28}$$

To solve

$$h(s_0) = 0 \tag{5.6.29}$$

We have

$$s_0 = \exp\left(-\frac{\beta}{l} \left(H(\vec{p}', \vec{r}) + \frac{\vec{p}_s^2}{2Q} - E\right)\right) \tag{5.6.30}$$

and

$$h'(s) = \frac{dh}{ds} = \frac{l}{\beta s} \tag{5.6.31}$$

$$\delta[h(s)] = \frac{\delta(s - s_0)}{|h'(s_0)|} \tag{5.6.32}$$

We can rewrite (6.26) to be

$$\begin{aligned}
Z_{\text{Nose}} &= \frac{1}{N!} \int dp_s ds d\vec{p}'^N d\vec{r}^N \frac{\beta s^{3N+1}}{l} \delta\left[s - \exp\left(-\frac{\beta}{l} \left(H(p', r) + \frac{p_s^2}{2Q} - E\right)\right)\right] \\
&= \frac{1}{N!} \frac{\beta}{l} \exp\left(\frac{3N+1}{l} E\right) \int dp_s \exp\left(-\beta \frac{3N+1}{l} \frac{p_s^2}{2Q}\right) \int d\vec{p}'^N d\vec{r}^N \exp\left(-\beta \frac{3N+1}{l} H(p', r)\right) \\
&= \frac{C}{N!} \int d\vec{p}'^N d\vec{r}^N \exp\left(-\beta \frac{3N+1}{l} H(p', r)\right)
\end{aligned} \tag{5.6.33}$$

If we select  $l = 3N + 1$ , we have

$$\begin{aligned}
 \bar{A} &= \lim_{\tau \rightarrow \infty} \int_0^{\tau} dt A\left(\frac{\vec{p}(t)}{s(t)}, \vec{r}(t)\right) \equiv \left\langle A\left(\frac{\vec{p}}{s}, \vec{r}\right) \right\rangle_{\text{Nose}} \\
 &= \frac{\int dp'^N dr^N A(\vec{p}', \vec{r}) \exp(-\beta H(\vec{p}', \vec{r}))}{\int dp'^N dr^N \exp(-\beta H(\vec{p}', \vec{r}))} \\
 &= \langle A(\vec{p}', \vec{r}) \rangle_{NVT}
 \end{aligned} \tag{5.6.34}$$

It should be noted that there is a certain transformation:

$$\begin{aligned}
 r' &= r \\
 p' &= p/s \\
 s' &= s \\
 \Delta t' &= \Delta t/s
 \end{aligned} \tag{5.6.35}$$

If we are to sample in a real time interval, it can be proven that we can select  $l = 3N$ , and the expanded Hamiltonian is a conserved quantity:

$$H_{\text{Nose}} = \sum_{i=1}^N \frac{\vec{p}_i'^2}{2m_i} + E_p(\vec{r}'^N) + \frac{s'^2 p_s'^2}{2Q} + \frac{l}{\beta} \ln s' \tag{5.6.36}$$

The equations of motion becomes:

$$\begin{aligned}
 \frac{d\vec{r}'_i}{dt'} &= s \frac{d\vec{r}_i}{dt} = \frac{\vec{p}_i}{m_i s} = \frac{\vec{p}'_i}{m_i} \\
 \frac{1}{s} \frac{ds'}{dt'} &= \frac{s' p'_s}{Q} \\
 \frac{d\vec{p}'_i}{dt'} &= s \frac{d(\vec{p}_i/s)}{dt} = \frac{d\vec{p}_i}{dt} - \frac{1}{s} \vec{p}_i \frac{ds}{dt} = -\frac{\partial E_p(\vec{r}'^N)}{\partial \vec{r}'} - \left(\frac{s' p'_s}{Q}\right) \vec{p}_i \\
 \frac{d}{dt'} \left( \frac{s' p'_s}{Q} \right) &= \frac{s}{Q} \frac{dp_s}{dt} = \frac{\sum_i \frac{\vec{p}_i'^2}{m_i} - \frac{1}{\beta}}{Q}
 \end{aligned} \tag{5.6.37}$$

To implement the Nosé-Hoover algorithm, we can set

$$\zeta = \frac{s' p'_s}{Q} \tag{5.6.38}$$

so that (we drop all the primes)



$$\begin{aligned}
\frac{d\vec{r}_i}{dt} &= \frac{\vec{p}_i}{m_i} \\
\frac{1}{s} \frac{ds}{dt} &= \zeta \\
\frac{d\vec{p}_i}{dt} &= -\frac{\partial E_p(\vec{r}^N)}{\partial \vec{r}_i} - \zeta \vec{p}_i \\
\frac{d}{dt'} \zeta &= \frac{\sum_i \frac{\vec{p}_i^2}{m_i} - \frac{1}{\beta}}{Q} \\
H_{\text{Nosé}} &= \sum_{i=1}^N \frac{\vec{p}_i^2}{2m_i} + E_p(\vec{r}^N) + \frac{1}{2} \zeta Q + \frac{l}{\beta} \ln s, l = 3N
\end{aligned} \tag{5.6.39}$$

Another thing that we should keep in mind is that the standard Nosé-Hoover thermostat is appropriate to deal with systems with one conserved property or with a fixed center of mass where there is no external force. For a system with  $M$  degrees of freedom, the Nosé-Hoover chain should be introduced [20].

For systems simulated in the  $NPT$  ensemble where pressure should be kept a constant, there are other algorithms such as the Berendsen barostat or the Hoover barostat that should be implemented during the simulation. The basic concepts are similar to the thermostats that we have discussed, and the detailed explanation may be found in Refs. [14–20].

#### 5.6.2.4 Ewald Summation

To calculate long-range interactions with the PBC, the Ewald Summation [21] should be implemented for the reason that the long-range interactions cannot be directly truncated at a certain cutoff distance. The basic idea of the Ewald Summation is that each point charge is screened by a Gaussian-shaped charge cloud, so that the interactions are divided into two parts: short-ranged in the real space, and long-ranged in the real space which can be truncated after converted into the reciprocal space by Fourier transform.

The potential generated by a point charge with the PBC is

$$E_p = \frac{1}{2} \sum_{\vec{n}}' \sum_i^N \sum_j^N \frac{q_i q_j}{|\vec{r}_{ij} + \vec{n}|} \tag{5.6.40}$$

where  $q_i$  is the charge of ion  $i$ ,  $\vec{r}_{ij}$  is the distance between ions  $i$  and  $j$ , and  $\vec{n} = (n_x, n_y, n_z)$ . The applied charge distribution surrounding ion  $i$  is a Gaussian one:

$$\rho_i(r) = q_i \alpha^3 \exp(-\alpha^2 r^2) / \pi^{\frac{3}{2}} \tag{5.6.41}$$

By applying the Fourier Transform, we have:

$$E_p(\varepsilon_s = 0) = \frac{1}{2} \sum_i^N \sum_j^N \left[ \begin{aligned} & \sum_{\vec{n}=0}^{\infty} \frac{q_i q_j \operatorname{erfc}(\alpha |\vec{r}_{ij} + \vec{n}|)}{|\vec{r}_{ij} + \vec{n}|} \\ & + \frac{1}{\pi L^3} \sum_{k \neq 0} q_i q_j \frac{4\pi^2}{k^2} \exp\left(-\frac{k^2}{4\alpha^2} \cos(\vec{k} \cdot \vec{r}_{ij})\right) \\ & - \left(\frac{\alpha}{\sqrt{\pi}}\right) \sum_{i=1}^N q_i^2 + \frac{2\pi}{3L^2} \left| \sum_{i=1}^N q_i \vec{r}_i \right|^2 \end{aligned} \right] \quad (5.6.42)$$

where

$$\begin{aligned} \vec{k} &= \frac{2\pi}{L} (l_x, l_y, l_z) \\ \operatorname{erfc}(x) &= \frac{2}{\sqrt{\pi}} \int_x^{\infty} \exp(-t^2) dt \end{aligned} \quad (5.6.43)$$

The first term in Eq. (6.43) is the expression of interaction in real space, while the second term represents the one in the reciprocal space. The third term eliminated the self-interaction that is redundantly added in the first two terms whilst introducing the last term makes it the expression in vacuum.

The computational complexity of the Ewald Summation is around  $O(N^2)$  using a fixed cutoff and around  $O(N^{\frac{3}{2}})$  if we optimize the cutoff parameters dynamically. Alternatively, implementing the Particle Mesh Ewald (PME) [21] or Particle-Particle Particle-Mesh (PPPM) [22] algorithms can discretize the space to utilize the fast Fourier transform (FFT) algorithm, which reduces the computational complexity from  $O(N^2)$  to  $O(N \log N)$ .

### 5.6.2.5 Neighbor List

A method that can largely reduce computational complexity is the Neighbor List method which is introduced by Verlet, and is also referred to as the Verlet List method [23]. As we know, computational complexity for iterating all pairs to calculate the forces is  $O(N^2)$ , but many short-range forces with a distance larger than the cutoff distance can be effectively treated as zero. By introducing the neighbor list method, we can divide the simulation space into cubes whose side lengths equal to the cutoff distance. Only two particles in the nearest 27 cubes (including itself) are iterated when calculating the force, because the force between two particles not in the nearest cubes is always zero. The computational complexity is then reduced to  $O(N)$ .

## 5.7 Simple Data Analysis

### 5.7.1 Energy-Related Data

During the simulation process, we can observe and measure the instantaneous values of the system total energy  $E$ , the kinetic energy  $E_k$ , and the potential energy  $E_p$ . Thus, the instantaneous temperature can be calculated by

$$T = \frac{2}{d \cdot k_B} E_k \quad (5.7.1)$$

where  $d$  is the spatial dimension. The instantaneous pressure can be calculated as

$$P = \rho k_B T + \frac{1}{d \cdot V} \sum_{i < j} \vec{f}(\vec{r}_{ij}) \cdot \vec{r}_{ij} \quad (5.7.2)$$

The heat capacity in the  $NVT$  ensemble becomes

$$C_V^{NVT} = \frac{\langle E_p^2 \rangle - \langle E_p \rangle^2}{k_B T^2} + \frac{3}{2} N k_B \quad (5.7.3)$$

and the heat capacity in the  $NVE$  ensemble follows

$$\langle E_k^2 \rangle - \langle E_k \rangle^2 = \frac{3k_B^2 T^2}{2N} \left( 1 - \frac{3k_B}{2C_V^{NVE}} \right) \quad (5.7.4)$$

### 5.7.2 Correlation Coefficients

The correlation function between variables  $A$  and  $B$  is defined as:

$$C(A, B) = \langle AB \rangle - \langle A \rangle \langle B \rangle \quad (5.7.5)$$

After normalization, Eq. (7.5) becomes

$$c(A, B) = \frac{\langle (A - \langle A \rangle)(B - \langle B \rangle) \rangle}{\sqrt{\langle (A - \langle A \rangle)^2 \rangle \langle (B - \langle B \rangle)^2 \rangle}}, \quad c \in [0, 1] \quad (5.7.6)$$

The *time correlation function* is

$$C(A, B, t) = \langle A(t_0)B(t_0 + t) \rangle$$

$$c(A, B, t) = \frac{\langle (A(t + t_0) - \langle A(t + t_0) \rangle)(B(t_0) - \langle B(t_0) \rangle) \rangle}{\sqrt{\langle (A(t + t_0) - \langle A(t + t_0) \rangle)^2 \rangle \langle (B(t_0) - \langle B(t_0) \rangle)^2 \rangle}} \quad (5.7.7)$$

whereas the *time autocorrelation function* is

$$C(A, t) = \langle A(t_0)A(t_0 + t) \rangle$$

$$c(A, t) = \frac{\langle (A(t_0 + t) - \langle A(t_0 + t) \rangle)(A(t_0) - \langle A(t_0) \rangle) \rangle}{\langle (A(t_0) - \langle A(t_0) \rangle)^2 \rangle} \quad (5.7.8)$$

Correlations and fluctuations are the essence of statistical mechanics, which make the world so rich and colorful. Therefore, most of statistical mechanics theories are about correlations and fluctuations, and correlation functions stand as important mathematical tools of statistical mechanics as well as necessary data analysis for molecular simulations.

### 5.7.3 Structural Properties

#### 5.7.3.1 Radial Distribution Function (RDF)

The radial distribution function (RDF) represents the probability of finding another particle with a distance of  $r$  relative to one particle:

$$g(\vec{r}_1, \vec{r}_2) = \frac{N(N-1)}{\rho^2 Z_{NVT}} \int d\vec{r}_3 d\vec{r}_4 \dots d\vec{r}_N \exp[-\beta E_p(\vec{r}_1, \vec{r}_2 \dots \vec{r}_N)] \quad (5.7.9)$$

which is equivalent to

$$g(r) = \frac{V}{N^2} \left\langle \sum_i \sum_{j \neq i} \delta(r - r_{ij}) \right\rangle = \frac{\sum_i \sum_j \delta(r - r_{ij})}{M \cdot N \cdot \rho \cdot \frac{4}{3}\pi [(r + \Delta r)^3 - r^3]} \quad (5.7.10)$$

where  $M$  is the number of configurations being sampled,  $N$  is the number of particles, and  $\rho = \frac{N}{V} = \frac{N}{L^3}$ . The RDF for an ideal gas equals to one for all  $r$  because particles have no spatial correlations in an ideal gas.

### 5.7.3.2 Structure Factor

The RDF cannot be directly measured by experiment. However, we can introduce the Fourier transform of it in  $k$ -space where  $\vec{k} = \frac{2\pi}{L}(k_x, k_y, k_z)$ , which is the experimentally measurable *structure factor*, that is

$$S(k) = 1 + 4\pi\rho \int_0^{\infty} r^2 \frac{\sin kr}{kr} g(r) dr \quad (5.7.11)$$

### 5.7.4 Diffusion Coefficient

Apart from the structural properties, to understand the system dynamics, it is also important to investigate the diffusion coefficients of particles. The *mean square displacement* (MSD) marks the square of the average distance moved in time duration  $t$ . The MSD is a time autocorrelation function, based on which the *diffusion coefficient*  $D$  can be calculated from the Einstein Relation:

$$\langle \vec{r}^2(t) \rangle \equiv \langle \Delta \vec{r}^2(t) \rangle = \frac{1}{N} \sum_{i=1}^N \Delta \vec{r}_i^2(t) = 2dDt \quad (5.7.12)$$

where the parameter  $d$  labels the dimension of the system. Equation (7.12) is strictly correct only when time  $t \rightarrow \infty$ .

The diffusion coefficient can also be represented by time autocorrelation function of velocity. The MSD can be rewritten as the integral of velocity autocorrelation function:

$$\begin{aligned} \langle \vec{r}^2(t) \rangle &= \left\langle \left( \int_0^t dt' \vec{v}(t') \right)^2 \right\rangle \\ &= \left\langle \int_0^t \int_0^t dt' dt'' \langle \vec{v}(t') \vec{v}(t'') \rangle \right\rangle \\ &= 2 \int_0^t \int_0^{t'} dt' dt'' \langle \vec{v}(t') \vec{v}(t'') \rangle \end{aligned} \quad (5.7.13)$$

Also, we have:

$$\langle \vec{v}(t') \cdot \vec{v}(t'') \rangle = \langle \vec{v}(t' - t'') \cdot \vec{v}(0) \rangle \quad (5.7.14)$$

Thus, according to Eq. (7.13),

$$\begin{aligned}
 2dD &= \lim_{t \rightarrow \infty} 2 \int_0^t dt'' \langle \vec{v}(t' - t'') \cdot \vec{v}(0) \rangle \\
 \Rightarrow D &= \frac{1}{d} \int_0^\infty dt \langle \vec{v}(t) \cdot \vec{v}(0) \rangle
 \end{aligned} \tag{5.7.15}$$

The term  $\langle \vec{v}(t) \cdot \vec{v}(0) \rangle$  is referred as velocity autocorrelation function (VACF). Equation (7.15) is a *Green-Kubo relation*, which relates the time autocorrelation function of microscopic properties to macroscopic thermodynamic properties. There are few other examples of the Green-Kubo relations:

1. Shear viscosity:

$$\begin{aligned}
 \eta &= \frac{1}{Vk_B T} \int_0^\infty \langle \sigma^{xy}(0) \sigma^{xy}(t) \rangle dt \\
 \sigma^{xy} &= \sum_{i=1}^N \left( m_i v_i^x v_i^y + \frac{1}{2} \sum_{i \neq j} x_{ij} f_y(r_{ij}) \right)
 \end{aligned} \tag{5.7.16}$$

2. Thermal conductivity:

$$\begin{aligned}
 \lambda_T &= \frac{1}{Vk_B T^2} \int_0^\infty \langle j_z^e(0) j_z^e(t) \rangle dt \\
 j_z^e &= \frac{d}{dt} \sum_{i=1}^N \frac{z_i}{2} \left( m_i v_i^2 + \sum_{i \neq j} v(r_{ij}) \right)
 \end{aligned} \tag{5.7.17}$$

3. Electric conductivity:

$$\begin{aligned}
 \sigma_e &= \frac{1}{Vk_B T} \int_0^\infty \langle j_x^{el}(0) j_x^{el}(t) \rangle dt \\
 j_x^{el} &= \sum_{i=1}^N q_i v_i^x
 \end{aligned} \tag{5.7.18}$$

## References

1. Hohenberg, P., Kohn, W.: Inhomogeneous electron gas. *Phys. Rev.* **136**(3B), 864–871 (1964)
2. Levy, M.: Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the  $v$ -representability problem. *Proc. Natl. Acad. Sci. U.S.A.* **76**(12), 6062–6065 (1979)
3. Kohn, W., Sham, L.J.: Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**(4A), A1133–A1138 (1965)
4. Noid, W.G., Liu, P., Wang, Y., Chu, J.W., Ayton, G.S., Izvekov, S., et al.: The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models. *J. Chemical Phys.* **128** (24), 244115 (2008)
5. Wang, Y., Noid, W.G., Liu, P., Voth, G.A.: Effective force coarse-graining. *Phys. Chem. Chemical Phys.* **11**(12), 2002–2015 (2009)
6. Marrink, S.J., Risselada, H.J., Yefimov, S., Tieleman, D.P., De Vries, A.H.: The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111**(27), 7812–7824 (2007)
7. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**(6), 1087–1092 (1953)
8. Frenkel, D., Smit, B.: *Understanding molecular simulation: from algorithms to applications*, 2nd edn. Academic Press, Inc. (1996)
9. Wood, W.W.: Monte Carlo calculations for hard disks in the isothermal-isobaric ensemble. *J. Chem. Phys.* **48**(1), 415–434 (1968)
10. McDonald, I.R.: NpT-ensemble Monte Carlo calculations for binary liquid mixtures. *Mol. Phys.* **23**(1), 41–58 (2002)
11. Najafabadi, R., Yip, S.: Observation of finite-temperature brain transformation (f.c.c.  $\rightarrow$  r b.c.c.) in Monte Carlo simulation of iron. *Scr. Metall.* **17** (10), 1199–204 (1983)
12. Allen, M.P., Tildesley, D.J.: *Computer simulation of liquids*, 2nd edn. Oxford University Press (2017)
13. Berendsen, H.J.C., Gunsteren, W.F.V.: *Molecular dynamics simulation of statistical mechanical systems*. North Holland Publishing Co Amsterdam (1986)
14. Berendsen, H.J.C., Postma, J.P.M., Van Gunsteren, W.F., Dinola, A., Haak, J.R.: Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**(8), 3684–3690 (1984)
15. Andersen, H.C.: Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **72**(4), 2384–2393 (1980)
16. Hoover, W.G.: Constant-pressure equations of motion. *Phys. Rev. A* **34**(3), 2499–2500 (1986)
17. Hoover, W.G.: Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **31**(3), 1695–1697 (1985)
18. Nosé, S.: A unified formulation of the constant temperature molecular dynamics method. *J. Chem. Phys.* **81**(1), 511–519 (1984)
19. Nosé, S.: A molecular-dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52** (2), 255–68 (1984)
20. Tuckerman, M., Berne, B.J., Martyna, G.J.: Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **97**(3), 1990–2001 (1992)
21. Darden, T., York, D., Pedersen, L.: Particle Mesh Ewald: An Nlog(N) Method for Ewald sums in large systems. *J. Chem. Phys.* **98**(12), 10089–10092 (1993)
22. Eastwood, J.W., Hockney, R.W.: Shaping the force law in two-dimensional particle-mesh models. *J. Comput. Phys.* **16**(4), 342–359 (1974)
23. Verlet, L.: Computer “Experiments” on classical fluids. I. thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* **159** (1), 98–103 (1967)

# Chapter 6

## Cocoon Silk: From Mesoscopic Materials Design to Engineering Principles and Applications



Wu Qiu and Xiang-Yang Liu

**Abstract** In this chapter, we present a comprehensive review on five levels of hierarchical structures of silk fibroin (SF) materials in correlation with macroscopic properties/performance such as the toughness, strain-stiffening, etc. It follows that the crystalline binding force turns out to be very crucial in the stabilization of the mesoscopic hierarchical structures of silk materials. In addition,  $\beta$ -crystallites,  $\beta$ -crystallite networks (nanofbrils) and the interactions between helical nanofbrils within nanofibril networks are determined to be the most essential meso structures, which greatly determine the macroscopic performance of various forms of silk materials. In this context, the characteristic structural factors such as the orientation, size, and density of  $\beta$ -crystallites are very relevant. It reveals that the formation of these structural elements is mainly controlled by the continuous intermolecular nucleation of  $\beta$ -crystallites. Consequently, the rational design and reconstruction of silk materials can be implemented by controlling the molecular nucleation via proper seeding (i.e., with carbon nanotubes). Moreover, by adding functional seeds, i.e., carbon nanotubes, one can reconstruct the mesoscopic networks of SF materials, which can endow silk materials with new performance, i.e., electrical conductivity. In general, the knowledge of the correlation between hierarchical structures and performance provides an understanding of the structural reasons behind the fascinating behaviors of silk materials. In this chapter, we also provide a summary on the state-of-art technologies in characterization of the structures of various levels of SF materials. The principles of SF nucleation and the application to reconstruct the meso structures of SF materials are also presented, together and the methods of investigation.

**Keywords** Hierarchical structure · Mesoscopic engineering · Functionalization · Silk fibroin materials · Nucleation model

---

W. Qiu

Department of Physics, National University of Singapore, 2 Science Drive 3, Singapore 117542, Singapore

X.-Y. Liu (✉)

Research Institution for Biomimetics and Soft Matter, Xiamen University, Lingfeng Building, Xiamen 31005, China

e-mail: [liuxy@xmu.edu.cn](mailto:liuxy@xmu.edu.cn)

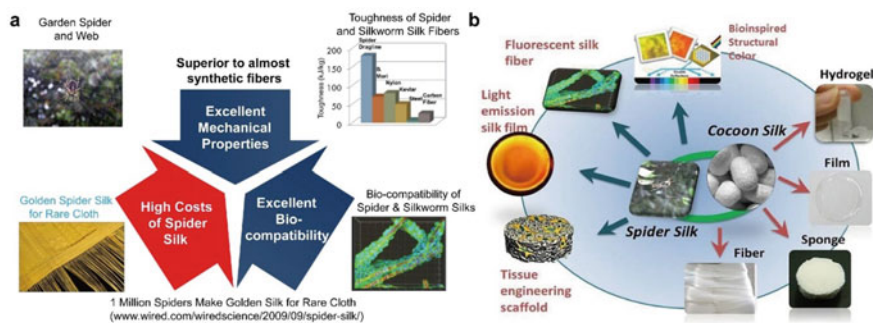


## 6.1 Introduction

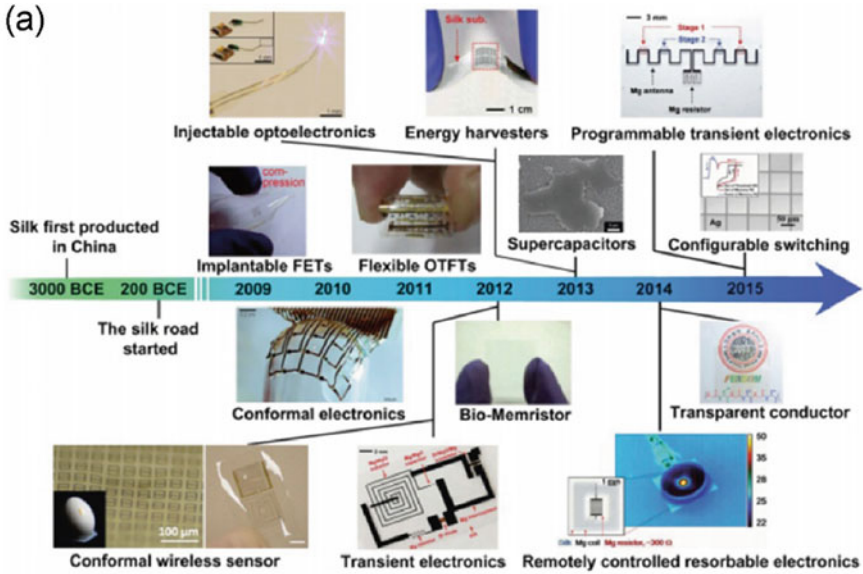
### 6.1.1 Silkworm Silk and Silk Fibroin (SF) Materials

Silk is a fine continuous protein fiber produced by spiders or insect larvae. Spiders spin threads of silk to make cocoons for their eggs and in their webs to catch prey, and such silks have been admired for their extraordinary mechanical properties since ancient times. However, while the application of spiders for silk production is limited because spiders cannot be raised, silk produced by insect larvae such as silkworms is a world-renowned fiber. Among the types of silk fibers, those spun by *Bombyx mori* (*B. mori*) silkworms are usually regarded as “the queen of fibers,” have been widely used in the textile industry for more than five thousand years, and are also symbols of the foreign trade of ancient China [1, 2]. With the development of materials science and a deeper investigation of silk, silkworm silk is no longer limited to the traditional textile industry, but is currently regarded as one of the most promising materials of the twenty-first century. Owing to its properties, such as excellent optical and electronic characteristics [3–5], robust mechanical properties [6–9], in vitro and in vivo biocompatibility [10], and slow proteolytic biodegradation [1], silk has attracted considerable interest within the scientific community (Fig. 6.1a) [11] and has been applied in areas such as tissue engineering, flexible electronics (Fig. 6.2a) [12], drug delivery [13], biological analysis [14], and lithography [15].

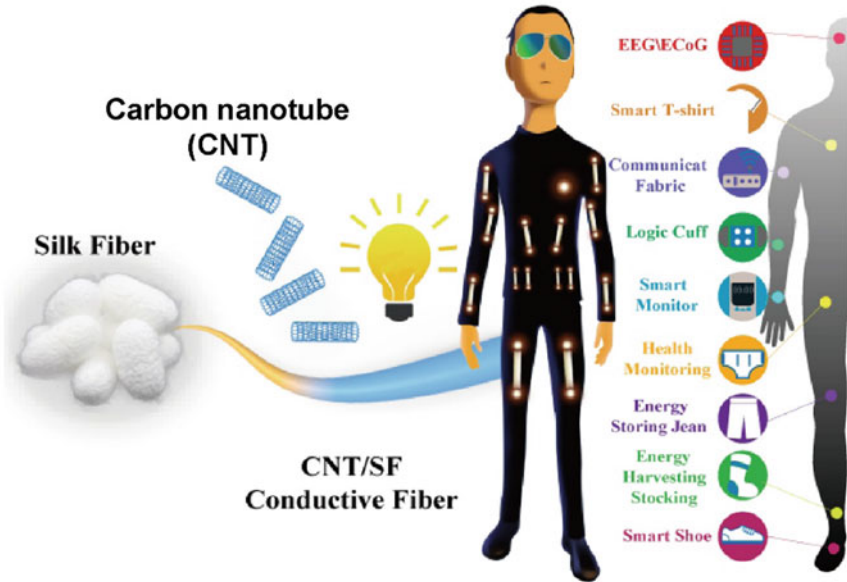
Furthermore, while functionalized silk materials exhibit various targetable optical and electronic characteristics and are versatile for functionalization, researchers have



**Fig. 6.1** **a** Spider silk and silkworm silk fibrous materials are well known for their unique properties, *i.e.*, excellent biocompatibility, heat conductivity, electrical insulating capacity, controllable dissolution, and wide optical window (260–2600 nm). Nevertheless, their relatively high cost greatly limits their application. **b** Various forms of silkworm cocoon silk and spider silk. Both spider and silkworm silk materials and their applications: silk hydrogels, films, and sponges, used in tissue engineering and biocompatible, optical, and electronic flexible devices. Reproduced with permission [16]. Copyright 2015, Royal Society of Chemistry



(b) Silk Fibroin Biocompatible Flexible Electronic Devices



**Fig. 6.2** a A brief timeline of the silk “electronic road.” Reproduced with permission [3]. Copyright 2016, Wiley-VCH. b Through carbon nanotube meso-fibrous reconstruction, a new conceptual silk meso-fibrous material has been developed for biocompatible electronic applications including electronic humidity sensors, remote respiratory condition monitoring, and diagnosis. Reproduced with permission [17]. Copyright 2020, Wiley-VCH

found that silk fibroin (SF) materials have potential in multiple applications, particularly for flexible optical and electronic devices [4]. For instance, the intrinsic performance as well as functionalization (by incorporating organic molecules or inorganic nanoparticles into the SF matrix) of SF materials make them ideal candidates for basic optical elements, light energy conversion devices, photochemical reactions, sensors, and bioimaging [4]. In addition, the programmable degradation behaviors of SF materials also enable the fabrication of a variety of SF-based flexible electronics for versatile applications including biomedical monitoring, therapy, biosensing, and memory devices [3, 5, 17].

### 6.1.2 *Structural Factors Correlated to Macroscopic Properties*

For many soft materials, including SF materials, the macroscopic performance is controlled largely by the unique structures, whose size ranges from the nanoscale to the mesoscale level (Fig. 6.2b). In general, the performance of soft materials is highly correlated to the following four structural factors: [16, 18, 19]

- (1) *Topology*: It demonstrates how joints/points are associated with each other. In principle, network modifications often begin with an alternation in topology.
- (2) *Correlation length,  $\xi$* : It describes the average distance between two correlated joints/points. In terms of hierarchical structure, the joints/points refer to structural units at the same level.
- (3) *Ordering/symmetry of structural units*: In many situations, structural units in a network are anisotropic. The ordering/symmetry enable us to understand how these structural units are patterned.
- (4) *Strength of interactions*: Notably, the interactions between adjacent structural units can be of the physical, thermodynamic, or chemical type.

The stability of networks is largely controlled by the strength of these interactions. To obtain a comprehensive understanding of the correlation between the structure and performance of soft materials, a more precise insight into the hierarchical structures of soft materials is required; this has also inspired scientists and engineers to fabricate advanced functionalized soft materials, by effectively manipulating the aforementioned structural factors [20]. Although numerous attempts have been made to characterize the hierarchical structures of silk materials, the mechanism by which the multi-level structures of silk materials are constructed and the corresponding structural factors affecting macroscopic performance remain debatable.

### 6.1.3 Scope of This Chapter

This chapter attempts to address the following questions: What are the hierarchical structures constituting SF materials? How do different levels of structures synergistically affect the macroscopic properties? Which techniques can be applied for structural characterization? This chapter will provide a fundamental, refined description of the hierarchical network structures of SF materials in a bottom-up manner, which will enable us to comprehensively investigate the above-mentioned topics. Moreover, the correlation between structure and performance is discussed accordingly. In Sect. 6.2, we present a brief overview of the history of sericulture in China. In addition, the life cycle of silkworms as well as the silk fabrication process are also included. In Sect. 6.3, we will first investigate the different levels of the hierarchical structures of SF materials and discuss how they can be correlated to mechanical strength. Then, in Sect. 6.4, we will introduce several structural characterization and imaging technologies. In Sect. 6.5, to illustrate how SF materials are self-assembled, we will put forward a nucleation-crystal growth model based on comprehensive observations and detailed experimental verification.

## 6.2 Silk Road: From Ancient Textiles to Smart Flexible Functional Materials

### 6.2.1 History and Life Cycle of Silkworms

The oldest textile fiber, silk, according to Chinese folklore, was first discovered by Empress His-Ling-Shi (the wife of the mythical Yellow Emperor) in approximately 3000 BC, while she was sipping her tea. The history of sericulture in China is very long (Fig. 6.3a and b) and is believed to have started during the Chinese Neolithic period. For instance, the oldest silk found in Henan Province, China, dates to circa 3630 BC. Other examples of very old silk in China are a group of silk threads, a braided silk belt, and a woven silk cloth fragment excavated from the Liangzhu culture site in Zhejiang Province, China, all of which date to approximately 2570 BC. Following the discovery of silk, it soon became a valued commodity in ancient China. With the development of sericulture, silk has become highly prized in China and throughout the world. For instance, the techniques of silkworm raising and silk fabrication were spread to nearby countries such as India and Japan (Fig. 6.3c). As time progressed to 139 BC (the Han dynasty), the ancient Chinese built a trade route stretching from Eastern China to the Mediterranean Sea called the Silk Road for the lucrative silk trade (Fig. 6.3d) [21].

In terms of silkworm species, it is generally accepted that silkworms can be classified into two main categories: mulberry and non-mulberry, based on their diet. Specifically, the mulberry silkworm refers to *B. mori* silkworms only. Given that *B. mori* silkworms are easily domesticated, they are also called domesticated silkworms.



**Fig. 6.3** Ancient Chinese sericulture and the silk road. **a** Long history of sericulture in ancient China (adapted from: <https://www.visiontimes.com/2017/05/20/silk-weaving-one-of-ancient-chinas-greatest-inventions.html>). **b** Image of silk weaving in ancient China (adapted from: <https://www.mariannsilk.com/pages/the-silk-story>). **c** Image of silk weaving in ancient Japan (adapted from: <https://www.pinterest.com/innakulagina/history-of-silk/>). **d** The route of famous Silk Road (adapted from: <http://www.italiamedievale.org/portale/la-via-della-seta-nel-tratto-transcaucasico-e-transcaspiano/?lang=zh>)

In comparison, non-mulberry silkworms are wild heterogeneous in nature and widely distributed throughout the world. Based on their location, non-mulberry silkworms can be further classified as temperate and tropical. Among them, typical temperate silkworms include *Antheraea pernyi* (*A. pernyi*) found in China, *Antheraea Assama* found in Japan, and *Attacus cynthia ricini* found in northeastern India; tropical non-mulberry silkworms mainly refer to *Antheraea mylitta* (*A. mylitta*), which is mostly confined to central India [21]. In terms of structure and properties, the silk spun by non-mulberry and mulberry silkworms are very similar. However, as non-mulberry silkworms mainly live in the wild, their cocoons are always contaminated by several inorganic minerals, which consequently bring great difficulty in further purification, spinning, and the weaving process.

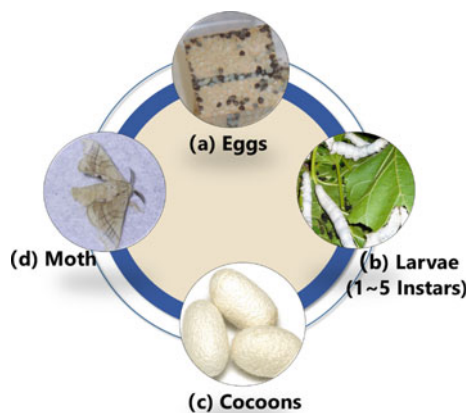
Although wild and domesticated silkworms differ greatly from each other in terms of their physiology, morphology, and feeding conditions, they still share substantial similarities in their life cycle. Taking *A. mylitta* and *B. mori* (cf. Fig. 6.4) as examples, the life cycle of both consists of four distinct stages: egg, larvae (five instars), pupa (inside cocoon), and adult moth. The duration of each stage varies by the species of silkworm and climatic conditions.

**Egg:** After copulation, the adult female moth usually lays 200–500 fertilized eggs on the surfaces of the leaves of their host plants under favorable conditions. The eggs are as small as ink dots.

**Larvae:** After hatching, 10–14 days pass, then the newly born larvae, which display tiny black hairs, begin to eat and grow continuously. The larvae pass through a total of five different molting phases, between which the interval is termed the “instar.” The duration of successive larval instars varies from three to eight days, whereas molting takes approximately 1 d. At the end of the fifth instar, the larvae stop eating, and their bodies turn translucent, indicating that they are ready to spin raw silk around themselves.

**Fig. 6.4** Four main stages of silkworms in the life cycle.

**a** Eggs and newly born larvae. **b** Larvae in the fifth instar. **c** Larvae make cocoons by spinning raw silk around themselves and metamorphose into pupa inside cocoons. **d** Moths break out of cocoons



**Pupa:** As indicated above, the larvae enclose themselves in a protective shell called a cocoon. Inside cocoons, the larvae are almost motionless until their metamorphosis into a pupa occurs, which usually takes two or three weeks.

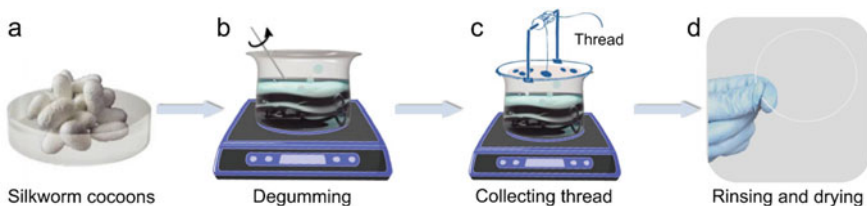
**Moth:** Once the conditions are favorable, the pupa releases proteolytic cocoonase, an enzyme that partially dissolves the cocoon walls so that the adult moth can emerge for mating. Male moths vary significantly from female moths in their body size as well as physiology; that is, the males are more active, and the females have a larger abdomen.

## 6.2.2 Processing and Fabrication of Cocoon Silk and Various Forms of SF Materials

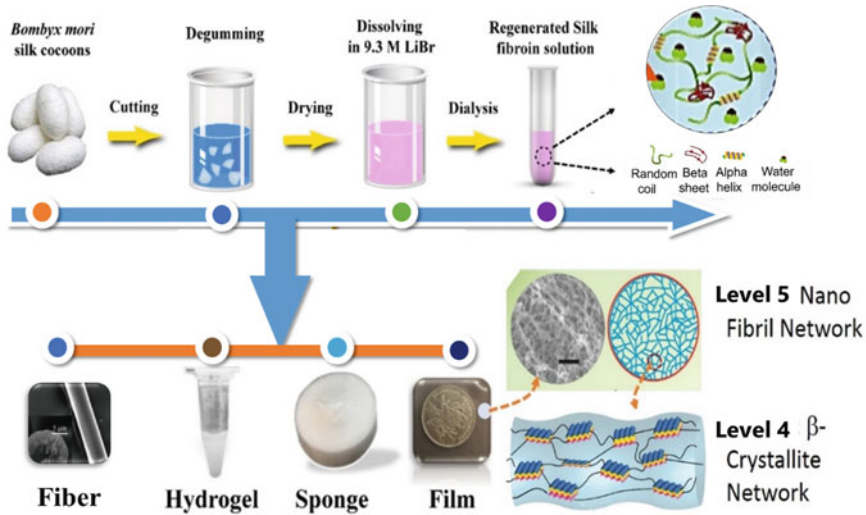
During cocoon formation, the silkworms swing themselves from side to side in a figure-eight and continuously produce silk from their silk glands. In general, an individual silkworm spins approximately 800–1600 meters of filaments [22]. Note that the quantity of usable quality silk filament in each cocoon is small, therefore, it usually takes thousands of silkworm cocoons to produce a kilogram of raw silk fibers.

In the typical fabrication process of silk from cocoons, the first step is to boil the cocoon in hot water (or sodium carbonate solution), which kills the silkworm pupa and removes the sericin coating (Fig. 6.5). This is known as the degumming process. Afterwards, a single silk filament can be gathered carefully by brushing the undamaged cocoon to find the outside end. To form a thread, many single filaments are combined and then drawn under tension through several guides and wound onto reels. The threads can be piled to form a yarn. After cleaning and drying, the raw silk can be further decorated with specific dyes to fabricate commercial colored silk cloth.

To prepare SF materials, the initial step is to prepare a large amount of regenerated silk fibroin (RSF) solution. A typical processing procedure was described by Liu et al. and is schematically shown in Fig. 6.6 [18]. After dialysis, the purified silk fibroin



**Fig. 6.5** Typical steps of silk processing. **a** Degumming: removing spoiled threads on outside of cocoons. **b** Boiling neat cocoons in hot alkaline solution. **c** Gathering single silk filament and fixing it into a collector. **d** Post treatments: rinsing and drying



**Fig. 6.6** Illustration of preparation of for the four different forms of regenerated SF materials: *fiber*, *hydrogel*, *sponge*, and *film*. **Levels 4 and 5** of network structures of SF materials are given

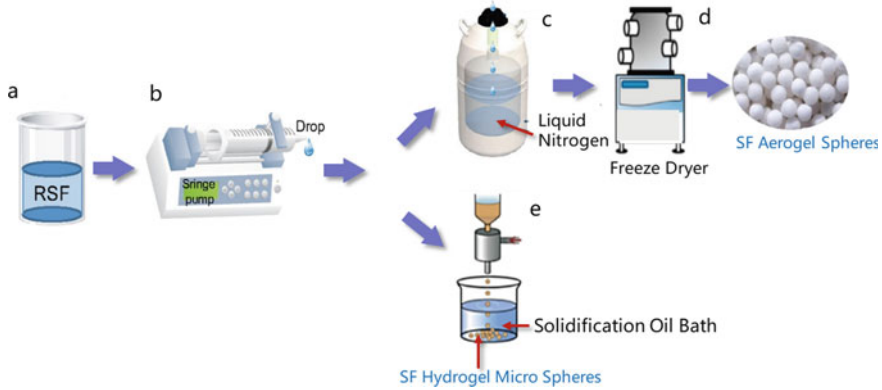
aqueous solution can serve as the basis for fabricating a variety of SF materials, including films, hydrogels, scaffolds/sponges, and artificial fibers.

Note that the four forms of SF materials displayed by Fig. 6.6 concern the bulk phase. In many cases, micro-spherical phases are of high relevance. They are of crucial applications in drug delivery and control release, tissue engineering, stem cells therapy and other biomedication etc. [23]. The SF micro spherical phases often include micro hydrogel spheres and micro aerogel (sponge) spheres [24]. The procedures of producing these microspherical phases are illustrated by Fig. 6.7.

### 6.3 Hierarchical Mesoscopic Network Structure of SF Materials in Correlation with Macroscopic Performance

Because the mechanical performance of silk fibers is largely determined by the unique hierarchical structures of the fibers, considerable efforts have been made over the past few years to investigate their structural secrets at the nanoscale and mesoscale levels, which form the basis of functioning of silk fibers at the macroscopic scale. The following structural models have been proposed and rapidly developed: [18, 26–29] (1) a semi-crystallite (bulk network) model based on polymer physics [28, 30]. This model considers silk fiber to be a composite material in which the crystalline regions are embedded into an amorphous matrix (made of rubber-like chains). Notably, the crystalline regions (mainly referred to as stiff  $\beta$ -sheet crystallites

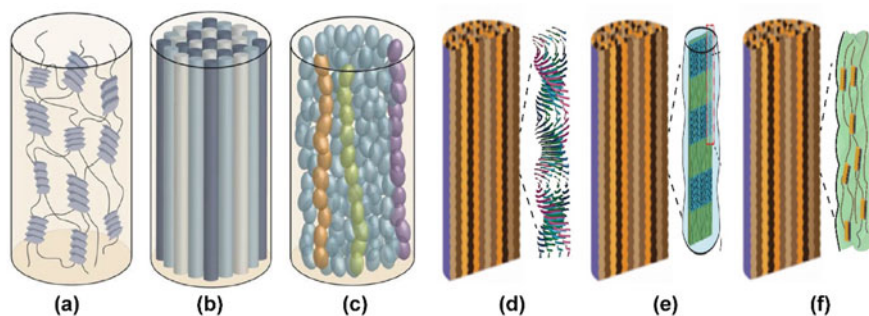




**Fig. 6.7** Shema of the procedures of producing RSF micro spheres. **a** RSF solutions prepared using traditional methods. **b, c** A RSF solution is transferred into a syringe which is mounted on a computer-controlled pump. **c** RSF solution is injected into liquid nitrogen with a constant speed and quenching into frozen spheres. **d** Upon freeze drying, water within RSF spheres is removed and SF aerogel spheres are obtained. **e** Solidification of RSF hydrogel spheres from RSF solutions. The solidification oil can be cooking oil + methanol, or ethanol, or acetone or a mixture [24, 25]

with a strong modulus) serve as multifunctional cross-links and cause silk fibers to have great strength, whereas the non-crystalline regions (amorphous matrices) are responsible for their excellent elasticity. To some degree, this model can successfully reproduce the complex stress–strain curves of silk fibers through molecular dynamics simulations that are consistent with experimental data. However, according to the simulation parameters, the modulus of these  $\beta$ -sheet crystallites was as high as 160 GPa; this value greatly exceeded the value observed during X-ray diffraction (XRD) experiments. In addition, this model is simple and only considers the nanoscale molecular level structures of silk fibers. (2) The cylindrical fibril model, which is based on the observation of fibrils (with a diameter ranging from 90 to 170 nm) on the surface of silkworm cocoon silk fibers [29]. Compared to the bulk network model, this model explains the mechanical performance of silk fibers from a meso-scale point of view. Owing to the interactions between adjacent fibrils, the loading stress can be efficiently dissipated; thus, the strength of the fibers is enhanced. (3) The micellar model [26, 31] claims that the fibril-like morphology of the fiber surface should arise from the coalescence and elongation of micellar structures. This model assumes that nano-globule micelles (with a diameter of  $\sim 100$  nm) are the basic micro-structural building units of silk fibers, instead of fibrils. Recently, these assumptions have been found to be debatable because of the latest findings that show the presence of numerous, significantly thinner ( $\sim 30$  to 50 nm), helically twisted nanofibrils along the fibrous axis of the fiber surface [18, 32–34].

As discussed earlier, the outstanding mechanical performance of silk fibers is synergistically determined by their different structural levels. However, all the models discussed above provide incomplete information. Because silk fibers consist of numerous crystallites, complete information regarding the structure is essential for



**Fig. 6.8** Several models describing possible mesoscopic structure of natural spun silk. **a** Bulk network model, **b** cylindrical fibril model, **c** micellar model, **d** amyloid fibril-like model, **e** slab segment model, and **f** nano-fishnet model

obtaining further insight into how these crystallites are linked to crystal networks. Typically, at least three molecular crystal networks can exist, including: (4) a molecular network that is similar to that observed among amyloid fibrils (amyloid fibril-like model) [18, 35], (5) the more ordered slab-segment structure (the slab-segment model proposed by Oroudjev et al.) [36], and (6) the fishnet structure in which  $\beta$ -crystallites serve as crosslinkers [18]. The latest results have ruled out the possibility of both the amyloid fibril-like and slab-segment models [18]. Indeed, the strong  $\beta$ -crystallites serving as the nodes in the nano-fishnet can reinforce the silk fibers by sharing external forces within the optimized network. Therefore, the adoption of the fishnet topology in crystal networks is a natural selection process for obtaining silk fibers exhibiting excellent macroscopic mechanical performance. Details regarding models with different crystal network topologies are comprehensively introduced in the following (Fig. 6.8).

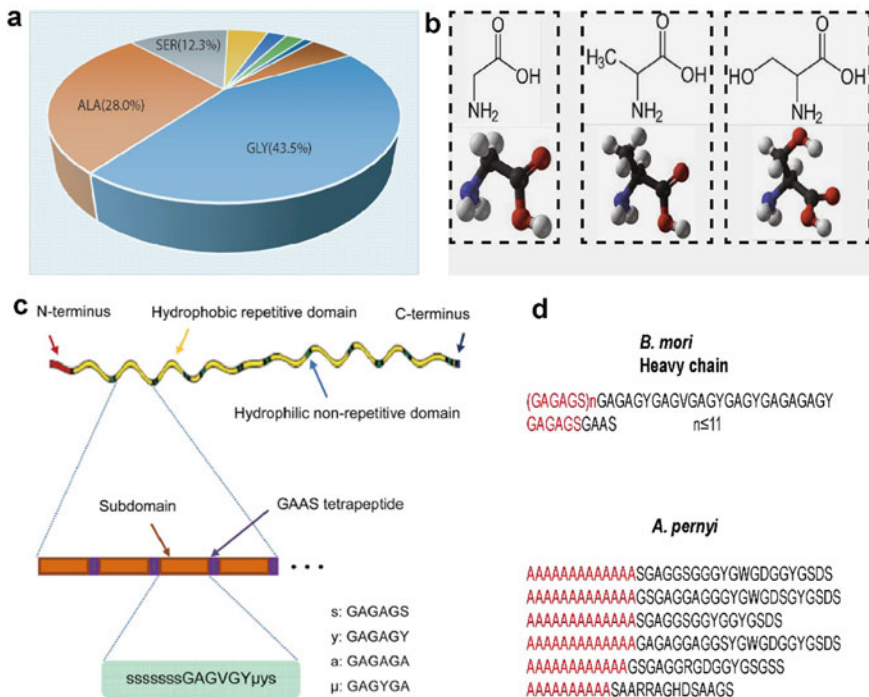
### 6.3.1 Primary (Level One) Structure of SF Materials

The primary structure of SF materials is defined as the amino acid sequence present within their protein molecules. Silkworm cocoon silk fibers are reported as mainly comprising two basic structural proteins: fibroin and sericin proteins. Notably, the sericin proteins located outside the fibroin proteins act as a glue that brings the two fibroin brims together. The sericin proteins are hydrophilic and can easily be removed by boiling the fibers in hot water. However, the fibroin proteins are mostly hydrophobic and can be further divided into two main categories—based on their molecular weight—of light (L) and heavy (H) fibroin chains with molecular weights of  $\sim 25$  and  $350$  kDa, respectively [22, 37]. The L and H chains are linked together via a single disulfide bond at the C-terminus and consequently form the H-L complex [38]. The H-chains of SF materials are generally responsible for the extraordinary mechanical performance of the fibers because they can self-assemble into discrete

$\beta$ -sheet crystallites ( $\beta$ -crystallites) [16, 18]. L-chains play a much less important role because their sizes are significantly smaller than that of H-chains; moreover, no L-chain amino acid sequences have been observed in the crystalline region.

The H- and L-chains are composed of 5263 and 266 amino acids, respectively. Specifically, the H-chain is primarily composed of the three simplest amino acids, glycine (G) (~43.5%), alanine (A) (~28%), and serine (S) (~12.3%). In addition, nearly 5% of tyrosine (Y) is also observed to be present (cf. Fig. 6.9a) [22, 37, 39]. Apart from the above four amino acids, the next most abundant amino acids include valine (V), aspartic acid (D), phenylalanine (F), glutamic acid (E), threonine (T), and isoleucine (I); however, in total, all these amino acids are present only in small amounts, that is, less than 2% [22].

In terms of amino acid organization, the SF H-chain is determined to be a regular biopolymer; it consists of 12 large hydrophobic domains (also denoted repetitive domains, named R01–R12) that are interspersed with 11 smaller hydrophilic domains (denoted non-repetitive amorphous domains, named A01–A11) [22]. Further insight into these hydrophobic domains reveals that they are composed of dipeptide units in



**Fig. 6.9** Primary structure of silk materials. **a** Amino acid composition of fibroin molecules. **b** Illustration of three most abundant amino acids. **c** Amino acid sequence of a fibroin heavy chain. Reproduced with permission [22]. Copyright 2015, Elsevier. **d** Representative repetitive sequences of *B. mori* and *A. pernyi* silkworm silk fiber.  $\beta$ -sheet constructing units are denoted in red for visual guidance

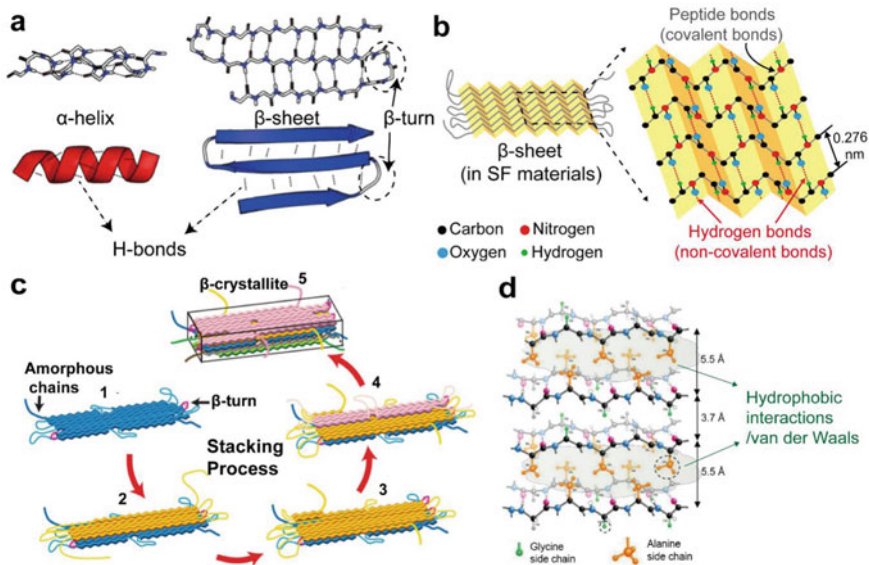
the form of glycine-X (GX), where X can be alanine, serine, tyrosine, valine, or threonine. Specifically, the hydrophobic domains can be divided into four distinct subdomains, including the GAGAGS repeats, GAGAGY repeats, GAAS tetrapeptides, and irregular sequences [40]. Statistically, the hexapeptide GAGAGS and GAGAGY repeats are the most abundant; they are present in 433 and 120 copies, respectively, accounting for 72% of all the repeat dipeptides.

It is reported that these four subdomains play distinct roles. For instance, the  $(\text{GAGAGS})_n$  repeats are the main units constituting  $\beta$ -sheets [41] and are the only units that can organize themselves into well-defined  $\beta$ -sheet crystallites ( $\beta$ -crystallites). However, no well-defined crystalline structures have yet been formed by GAGAGY repeats. Although some evidence indicates that GAGAGY repeats might form  $\beta$ -crystallites, the structural units (especially the inter-sheet distances) within such  $\beta$ -crystallites are significantly different from those formed by  $(\text{GAGAGS})_n$  repeats. The GAAS tetrapeptides, which are located between the  $(\text{GAGAGS})_n$  repeats and serve as  $\beta$ -strand turns, can limit the size of  $(\text{GAGAGS})_n$   $\beta$ -crystallites. Similarly, by forming a loop structure with a distorted  $\Omega$  shape, irregular sequences can reverse the direction of  $\beta$ -strand backbones, which control the crystallite dimensions. In addition, a reversion of the backbone direction might also enable adjacent  $\beta$ -strands to be in contact with each other to facilitate the occurrence of inter-strand interactions.

Apart from domestic mulberry *B. mori* silkworms, many other arachnids or insects can also produce silk. Among these, the fibers from non-mulberry *A. mylitta* and *A. pernyi* silkworms are two typical examples and have been extensively studied. Irrespective of the sources, all these silk fibers are primarily composed of proteins and have a hierarchical structure similar to that of *B. mori* fibers. However, the amino acid sequences of these silk proteins can vary significantly from one species to another, which results in silk proteins having distinct mechanical properties. For instance, *A. pernyi* silk fibers reportedly present significant differences in both the amino acid compositions and amino acid sequences compared to those of *B. mori* silk fibers (Fig. 6.9d). The primary repetitive domains of *A. pernyi* silk fibers are the alternating tandem repeats of the poly-alanine motif. In addition, it was confirmed that such poly-alanine motifs serve as structural units for the construction of  $\beta$ -sheets and higher-level structures.

### 6.3.2 Secondary (Level Two) Structures of SF Materials

Secondary structure refers to the symmetrical structures formed by folding peptide chains, which are stabilized by hydrogen bonding (H-bonding) between the amine and carboxyl groups (the backbone-backbone and sidechain-sidechain hydrogen bonds are non-relevant). The most common secondary structural elements are  $\alpha$ -helices and  $\beta$ -sheets (Fig. 6.10a) [37]. Random coils refer to the multi-conformational state of peptides, unfolded proteins, and polypeptide chains. Sometimes, random coil is also adopted to describe unstructured conformations. It should be noted that random



**Fig. 6.10** Secondary and tertiary level of structures of silk materials. **a** Schematic of two typical secondary structures:  $\alpha$ -helix and  $\beta$ -sheet. **b** In  $\beta$ -sheets within silk materials,  $\beta$ -sheets adopt an anti-parallel arrangement. **c** Illustration of process of forming  $\beta$ -sheets from different molecules stacking into  $\beta$ -crystallite. **d** Inside  $\beta$ -crystallites, main forces maintaining the structure of  $\beta$ -crystallites by linking different  $\beta$ -sheet plates are the hydrophobic interactions/van der Waals interactions, especially those between alanine and serine. Reproduced with permission [22]. Copyright 2015, Elsevier

coils are not true secondary structures due to their irregularity.  $\beta$ -turns, which lead to a change in the direction of polypeptide chains, do not belong to a certain type of secondary structure by definition due to their structural irregularity [37]. Inside each  $\beta$ -turn, an individual hydrogen bond is formed between the backbone carbonyl oxygen of one residue (for instance, glycine in the sequence of GAAS) and the backbone amide of the residue three positions further along the chain (e.g., the serine in the GAAS sequence).

Concerning the  $\alpha$ -helix, the hydrogen bonds form between the oxygen atom of the C=O of each peptide bond. This occurs in the strand and the hydrogen atom of the N-H group of the peptide bond four amino acids below it in the helix. In this way, the direction of the H-bondings is roughly parallel to the  $\alpha$ -helix, and the hydrogen bonds help make this secondary structure especially stable [37]. In comparison, the H-bondings in the  $\beta$ -sheet are located between strands (inter-strand) rather than within strands (intra-strand). The carbonyl oxygen atoms in one strand of hydrogen bond with the amino hydrogen atoms of the adjacent strand (Fig. 6.10a) [37]. Similar to the  $\alpha$ -helix, this hydrogen bonding is the main interaction maintaining the structure of the  $\beta$ -sheet. It is noted that although individual H-bondings are relatively weaker than covalent bonds (e.g., 10–50 kJ/mol for H-bondings and 250 kJ/mol for covalent

bonds, respectively) [42], numerous H-bondings accumulated between strands can greatly reinforce the interactions and thus strengthen the  $\alpha$ -helix and  $\beta$ -sheet. In contrast, although adjacent  $\beta$ -strands can run either in the same direction (parallel arrangement) or in the opposite direction (antiparallel arrangement), the  $\beta$ -sheets in silk materials naturally choose an antiparallel  $\beta$ -strand arrangement (Fig. 6.10b) [37]. This is because the linear H-bonds (2.76 Å) within antiparallel  $\beta$ -strands are shorter than the non-linear H-bonds (2.97 Å) between parallel  $\beta$ -strands, which consequently results in greater  $\beta$ -sheet stability [22].

$\beta$ -sheets are structurally more compact and stable in aqueous environments and have stronger mechanical properties than random coils and  $\alpha$ -helices. In fact, most molecules within silk materials exhibiting  $\beta$ -sheet conformations (Silk II structure) are more likely to be insoluble. However, when the  $\alpha$ -helix conformation plays a dominant role (Silk I structure), such materials can be dissolved easily. In addition, because the  $\alpha$ -helix consists of easily movable chains, it is reasonable that a higher content of  $\alpha$ -helix within SF materials can lead to greater flexibility.

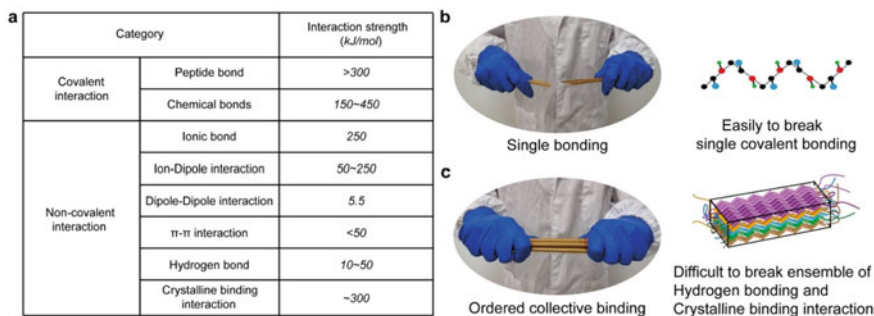
### 6.3.3 *Tertiary (Level Three) Structures of SF Materials and Crystalline Binding Interaction*

As mentioned at the beginning, SF materials, especially the silkworm cocoon fiber, display a very high toughness in terms of breaking energy. This can be attributed to the occurrence of  $\beta$ -crystallites and the crystal network. Typically, several secondary structural elements might be used to create a highly compact and organized three-dimensional (3D) protein structure, which is known as the tertiary structure. For SF materials, the tertiary structure can be defined as intermolecular  $\beta$ -crystallites (Fig. 6.10c) [16, 18]. Specifically,  $\beta$ -crystallites are comprised of several adjacent  $\beta$ -sheets from different molecules, among which hydrogen bonds and hydrophobic interactions/van der Waal's interactions play a key role in the  $\beta$ -crystallite (Fig. 6.5d) [16, 18, 22]. In this regard, these interactions can also be referred to as the crystalline binding interaction/force. From the point of view of crystallography,  $\beta$ -sheets and  $\beta$ -crystallites can be considered as a type of polymorphism: if  $\beta$ -sheets are considered as two-dimensional (2D) crystals,  $\beta$ -crystallites belong to a more stable form of 3D crystals. Although some  $\beta$ -sheets are capable of converting themselves into  $\beta$ -crystallites, it is not necessary for all  $\beta$ -sheets to be converted into  $\beta$ -crystallites. In general, several techniques, which will be introduced in the next section, can be applied to quantify the total  $\beta$ -conformation and  $\beta$ -crystallite content. The  $\beta$ -conformation content is always higher than the  $\beta$ -crystallite content because it consists of two types of polymorphism: the intramolecular 2D  $\beta$ -sheets and intermolecular 3D  $\beta$ -crystallites. The intramolecular  $\beta$ -sheet content can be calculated simply by deducting the  $\beta$ -crystallite content from the gross amount of  $\beta$ -conformation.

Recently, it has been widely accepted that different contents of  $\beta$ -sheets and  $\beta$ -crystallites give rise to distinct mechanical properties of different silk fibers. For instance, in *A. pernyi* silk fibers, the intramolecular  $\beta$ -sheet content is 23% [43]. In comparison, *B. mori* silk fibers exhibit a much smaller ratio of intramolecular  $\beta$ -sheet content (9% out of 49%) [33]. XRD analysis of the  $\beta$ -crystallite structure further confirms that the crystallites within the *B. mori* silkworm fibers exhibit an orthorhombic crystal lattice (with the space group of  $P22_12_1$ ) structure and have unit cell dimensions of  $a = 9.2$ ,  $b = 9.4$ , and  $c = 6.94$  Å, respectively, where  $a$  is the direction in which the  $\beta$ -sheets are stacked (inter-sheet distance),  $b$  is the direction of a  $\beta$ -sheet that is perpendicular to the strand axis (backbone-backbone hydrogen bonding direction, or inter-chain direction), and  $c$  is the direction of the strand axis [33].

Detailed investigation into the structure of  $\beta$ -sheets reveals that the side chains of adjacent  $\beta$ -strands within the same  $\beta$ -sheet plates are the same, that is, the front side of an entire  $\beta$ -sheet projects only the glycine side chains ( $-H$ ), whereas the back side projects only the alanine ( $-CH_3$ )/serine ( $-CH_2OH$ ) side chains. Above/beneath this  $\beta$ -sheet plate, the arrangement of the next plate is the opposite; thus, the  $\beta$ -sheet assembly is realized in a front-to-front and back-to-back manner. The side chains are arranged in such a way that the  $-CH_3$  and  $-CH_2OH$  groups of opposing  $\beta$ -sheets can be closely packed; this gives rise to inter-sheet interactions (Fig. 6.10d) [22].

Although both hydrogen bonding and hydrophobic interactions are much weaker than covalent bonds (Fig. 6.11a) [42]. Nevertheless, if the binding entities are in the form of a crystalline state, the binding force can be significantly enhanced. Unlike the non-crystalline state in which each non-covalent bond/interaction can be broken individually, in the crystalline binding state, the minimal breaking force of binding crystallites is equivalent to the simultaneous breakage of all the non-covalent binding interactions in a critical volume of crystallite [37]. This is due to the fact that the



**Fig. 6.11** Summary of variety of interactions involved in SF materials and enforcement strength in crystallization (*crystallization force*). **a** Table of interaction strength of different categories of interactions. **b** Illustration of strength enforcement due to  $\beta$ -crystallization, which is subject to collective enforcement by ensemble of hydrogen bonding between neighboring  $\beta$ -strands and crystalline binding force (hydrophobic interaction/van der Waals) between adjacent  $\beta$ -sheets within  $\beta$ -crystallites. Reproduced with permission [37]. Copyright 2019, Wiley-VCH

crystallites become unstable if the size of the crystallites is smaller than the critical size. This gives rise to the collectiveness of non-covalent bonds within the volume, displaying much stronger bonding. In this case, although a single noncovalent interaction is weak, the combination of noncovalent interactions can result in a strong bonding case. This encourages strong connections and stable binding points in the networks in silk materials [37]. In this regard,  $\beta$ -crystallization plays a key role in the stabilization of the meso structure of silk protein materials. Therefore, without  $\beta$ -crystallization, silk protein materials are very unstable and highly water soluble.

To acquire a comprehensive understanding of the role of hydrophobic interactions in maintaining the structural integrity of  $\beta$ -crystallites, the strength and number of hydrophobic interactions within different silk fibers were semi-quantitatively compared. According to the latest simulation results, the strength of the ensemble of hydrophobic interactions can be roughly estimated by counting the number of alanine and serine residues per sheet [18]. In addition, the regularity with which the alanine/serine pattern occurs is also relevant [18]. For instance, according to the protein sequence and residues forming  $\beta$ -sheets within *B. mori* cocoon silk (BMCS) and spider *Nephila antipodiana* eggcase silk (NAES) fibers, the number of times the alanine/serine pattern occurs within NAES is relatively low; consequently, the probability of the occurrence of hydrophobic interactions between  $\beta$ -sheets in NAES is smaller. Hence, the average peel-off force in BMCS fibrils is stronger than that for recombinant NAES fibrils ( $250 \pm 95$  pN and  $162 \pm 49$  pN, respectively) [18]. This difference can further explain why a higher breaking stress (550 MPa) is associated with natural BMCS fibers than with recombinant NAES fibers (256 MPa) [18].

Apart from the inter- $\beta$ -sheet interaction strength, the size of  $\beta$ -crystallites is also an important structural factor that influences the performance of silk materials. The size of  $\beta$ -crystallites in *B. mori* silk fibers and *A. pernyi* silk fibers was investigated using XRD and is shown in Table 6.1 [33, 37, 43]. To provide insights into the mechanical role of  $\beta$ -crystallites and the correlation between  $\beta$ -crystallite size and the mechanical properties of silk fibers, molecular modeling as well as simulations have been adopted. For instance, Wu et al. [44] proposed the  $\beta$ -sheet (within  $\beta$ -crystallites) splitting mechanism and found that, upon assuming that this mechanism is used, the

**Table 6.1** Structural parameters of silk fibers

Sample name	Overall content of $\beta$ -sheet (%)	Crystallinity (%)	Content of intramolecular $\beta$ -sheet (%)	Crystallite size (nm) <sup>a</sup>			References <sup>b</sup>
				<i>a</i>	<i>b</i>	<i>c</i>	
<i>B. mori</i> cocoon silk	49	40	9	2.3	4.1	10.3	33
<i>A. pernyi</i> silk	49	26	23	3.1	3.9	4.0	43

<sup>a</sup>Crystallite size along *a* (inter-sheet direction), *b* (inter-chain direction), and *c* (fiber axis) axes are measured at 0% strain for *B. mori* and *A. pernyi* silk

<sup>b</sup>The structural data are taken directly from references



predicted stress–strain profiles of silkworm silk fibers are in agreement with the results of experimental measurements. According to their reports, during the silk fiber stretching process, the extension linearly increases with the responsive force in the linear region, until the force reaches a threshold, after which the  $\beta$ -crystallites start to split. The splitting of  $\beta$ -sheets (within  $\beta$ -crystallites) results from the breakage of H-bondings, and assuming that the occurrence of H-bondings (along the  $\beta$ -strands) is uniform everywhere, a larger amount of force is required to extract longer molecules from  $\beta$ -crystallites because more H-bondings must be broken [44]. In other words, if the fibrous axis (the  $c$  direction) of a  $\beta$ -crystallite is larger, such crystallites become more robust. However, the beneficial effect attributable to crystallite size in the  $c$  direction (i.e.,  $L_c$ ) on the mechanical stability is limited. In contrast, Xu et al. [32] have shown that if the crystallite size  $L_c$  is larger than 6 nm, the influence of size on the splitting force becomes negligible. In natural silkworm silk fibers, the measured  $L_c$  values were more than 6 nm (Table 6.1). Thus, the splitting force of  $\beta$ -crystallites in silk materials should not be correlated to the crystallite size along the  $c$  direction [32]. However, Keten et al. [45] performed a series of large-scale molecular dynamics (MD) simulations to investigate the effects of crystallite size (in the  $b$  direction) and reported that small  $\beta$ -crystallites with a size of only a few nanometers can give rise to greater strength, stiffness, and toughness compared to large  $\beta$ -crystallites [45]. According to the results of their simulation study, in the crystal network, if the length of  $\beta$ -crystallites is 2–4 nm along the interchain direction ( $L_b$ ), the mechanical properties are optimal [45]. This experimental finding is in accordance with the reported results; the strength and toughness of silk fibers are enhanced when the  $\beta$ -crystallite size is reduced from 10 nm to 6.5 nm and further reduced to 3 nm. This can be attributed to the difference in geometry and stress distribution pathways of small and large  $\beta$ -crystallites. For large crystals, the hydrogen bonds are directly stretched because of tension (i.e., the  $\beta$ -crystallites are pulled in a direction parallel to that of the hydrogen bonds). In contrast, for small crystals that can be deformed when exposed to shear forces, hydrogen bonds are pulled orthogonal to the bonding direction. This shear deformation pathway can optimally utilize hydrogen bonds and consequently lead to a significant enhancement in the mechanical strength of  $\beta$ -crystallites [45].

From another perspective, compared with smaller crystallites, larger crystallites should contain more defects/mismatches, which also deteriorate their stability. According to crystallization theory, the formation of large crystallites normally relies on the package having a certain degree of ordering or symmetry [16]. During the normal crystal growth process, crystallites grow together via self-epitaxial nucleation, which consequently results in an ordered assembly. Nevertheless, under specific conditions, e.g., when the supersaturation is so high that the nucleation barrier for mismatched epitaxial nucleation decreases rapidly, some mismatches may occur between the parent crystals and nucleating layers of crystals. In this regard, the newly deposited layers of growing crystals would deviate from the optimal structural match position, and mismatched/misaligned structural packing would be observed [16].

### 6.3.4 *Level Four Structure of SF Materials: Fishnet-Like Crystallite Networks and Nanofibrils*

Merely crystalline binding is insufficient to construct tough materials. A tough material depends to a considerable extent on how these  $\beta$ -crystallites are connected. This is then associated with the fourth level of structure:  $\beta$ -crystallite networks. In general, the level four structures of a protein should refer to complexes composed of multiple subunits. The molecular  $\beta$ -crystallite networks of SF materials can be defined as a fourth-level structure [16, 18, 32]. Notably, each individual nanofibril is a crystallite networks in which amorphous chains link the crystallites together. In this regard, we can also regard the nanofibril (crystal network) as a level four structure. Nanofibrils are observed in all forms of SF materials (including fibers, hydrogels, films, and scaffolds), indicating that nanofibrils are the basic mesoscopic structural units of the hierarchical structure of SF materials.

In terms of the type of crystal network topology, recently, atomic force microscopy (AFM) has become a powerful tool for examining the nanostructures within nanofibrils by measuring the corresponding nanomechanical performance. Liu et al. [18] investigated the structure of nanofibrils obtained either from RSF solution or from natural SF fiber via AFM imaging, small angle X-ray scattering (SAXS), Fourier transform infrared (FTIR) spectroscopy, and XRD characterization and confirmed that these two types of nanofibrils share great similarities in structure and morphology. This similarity enables the use of individual RSF nanofibrils as an effective surrogate for natural SF fibers. After their initial investigation, they further used the regenerated nanofibrils to verify the type of topology of the  $\beta$ -crystallite network using AFM force spectroscopy [18]. By studying the unfolding force patterns of  $\beta$ -crystallites, Liu et al. presented a hypothesis about the mechanism by which the breakage of  $\beta$ -crystallites occurs during stretching and concluded that the  $\beta$ -crystallites are associated with each other in a nano-fishnet topology [18]. In the AFM force spectroscopy experiments, the AFM tip was used to pull out individual nanofibrils in order to probe the elasticity of  $\beta$ -crystallites. The manner in which the silk protein chains are connected via  $\beta$ -crystallites can significantly affect the dissipation of force from the AFM tip within the semi-crystalline networks; this is reflected by the measured force patterns. Typical sawtooth patterns were observed in the force versus extension trajectories of the regenerated nanofibrils. In these force patterns, the height of the force peaks corresponds to the strength of the hydrogen bonds between the  $\beta$ -strands in the  $\beta$ -sheet and/or the inter- $\beta$ -sheet interactions in the  $\beta$ -crystallites, while the changes in the level of extension between two adjacent peaks correspond to the released length of the polypeptides (contour length changes can be determined by fitting data to the worm-like chain model). The aforementioned sawtooth patterns of SF nanofibrils are distinct from the characteristic plateau force pattern observed for amyloid fibrils [35], indicating that the molecular network structures within amyloid fibrils and SF nanofibrils are different. This disparity between the amyloid fibrils and SF nanofibrils was also observed while examining the XRD

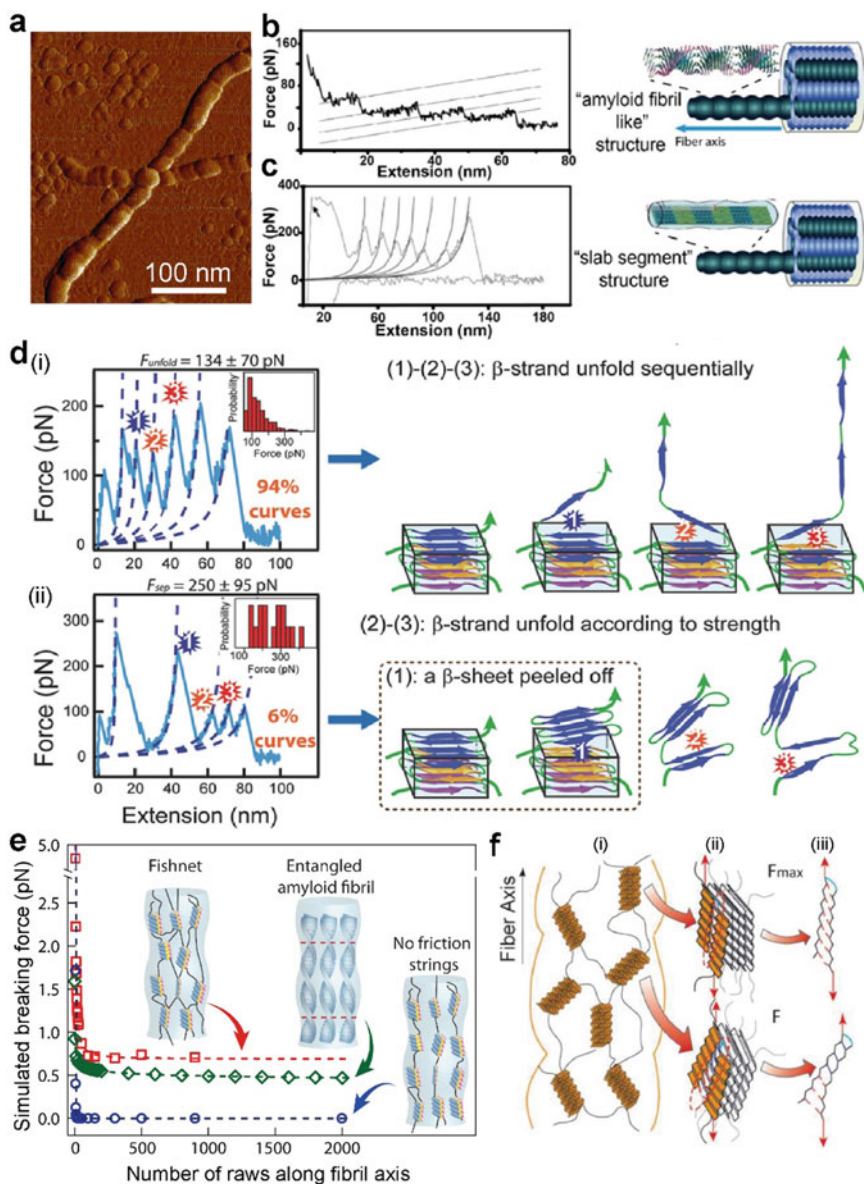
spectra; the  $\beta$ -strands in the crystallites were aligned along the fibrous axis in silk fibers but were perpendicular to the long axis of amyloid fibrils. Statistical results show that most force-extension curves (94%) of RSF nanofibrils exhibit sequential unfolding events involving random peak forces; however, no clear trend was observed. For the rest of the trajectories (6%), another characteristic pattern was identified [18]. The highest unfolding force peak was followed by a series of events, which shows a general upward trend in unfolding forces. Given the fact that hydrogen bonds between  $\beta$ -strands and hydrophobic interactions between  $\beta$ -sheet layers are responsible for the stability of  $\beta$ -sheets and  $\beta$ -crystallites, the two distinct trajectories imply two possible pathways for unfolding  $\beta$ -strands from  $\beta$ -crystallites. Pathway 1 involves the sequential unzipping of  $\beta$ -strands directly from  $\beta$ -crystallites (without affecting the inter- $\beta$ -sheet interactions), whereas Pathway 2 involves the peeling of a  $\beta$ -sheet off the crystallite first, followed by the unzipping of  $\beta$ -strands from the peeled-off  $\beta$ -sheet. In the case of Pathway 1, at each step of the unzipping of the  $\beta$ -strands, the force is mostly applied onto the strand directly connected to the AFM tip, while an equal amount of force is exerted on the other strands in the  $\beta$ -crystallites, owing to the integrated response of the  $\beta$ -crystallites. Hence, the strands directly linked to the AFM tip should experience a force much stronger than in other locations, so they are unzipped beforehand. This is then repeated for the next adjacent strand. As the strength of collective hydrogen bonding between different  $\beta$ -strands is different, the corresponding force pattern should result in peaks with different heights. According to the slab-segment model, the force distribution among  $\beta$ -strands in the  $\beta$ -sheet is similar to that in  $\beta$ -crystallites; hence, the force pattern should be similar to Pathway 1 [36]. However, if the  $\beta$ -crystallites are cross-linked with each other, as seen in a fishnet network, the peeled-out  $\beta$ -sheet will be stretched out between two anchor points, i.e., the AFM tip and the remaining  $\beta$ -crystallites. In this condition, an equal amount of stretching force is applied on all the  $\beta$ -strands within the peeled-off  $\beta$ -sheet because they are connected in a series. Thus, the weakest  $\beta$ -strands are unzipped first, followed by the others in the order of strength [18]. This can explain the existence of the 6% trajectories. Furthermore, Liu et al. [18] carried out a series of simulations to compare the relative strength of networks according to both pathways. The simulation results confirm that both the calculated possibility and the contour length change (of a  $\beta$ -sheet being peeled off from  $\beta$ -crystallites) according to the fishnet structure are in agreement with the AFM results [18]. This again indicates that the  $\beta$ -crystallites in the SF nanofibrils adopt a nano-fishnet topology. Furthermore, it is reported that a similar force pattern for recombinant spider eggcase silk fibrils has been observed, suggesting the prevalence of the molecular fishnet structure in animal silk nanofibrils of different types (Fig. 6.12e) [18]. It is noted that the adoption of the fishnet structure is the natural choice for obtaining outstanding toughness in the crystallite network. In fact, when using the Monte Carlo approach to simulate the breaking stress of different network topologies, it was calculated that the fishnet

structure was able to stabilize long silk fibers at a substantially reserved mechanical strength, which has greatly surpassed the breaking force predicted within the entangled amyloid fibril model and the no friction strings model (Fig. 6.12f) [18].

Several types of crystallite patterns can be observed within the framework of the nano-fishnet crystal network topology. For example, while  $\beta$ -crystallites can be effectively oriented adjacent to each other along the fibrous axis, they can also be oriented in a direction perpendicular to the fibrous axis. Recently, it has been determined that the nanofibrils in natural SF fibers are dominated by parallel  $\beta$ -sheets in which  $\beta$ -strands are parallel to the fibril axis. However, the cross- $\beta$ -sheet arrangement, in which  $\beta$ -strands are perpendicular to the fibril axis, can also exist in some non-fibrous SF materials [46]. Gong et al. [46] reported that SF materials can be selectively folded into  $\beta$ -sheets with either a cross- $\beta$ -sheet or parallel- $\beta$ -sheet arrangement by incubating SF solutions quiescently or under shear force, respectively. Considering the existence of shear forces during the natural spinning process of silk fibers, it is reasonable to assume that such shear forces play a dominant role in determining the arrangement of  $\beta$ -strands. Specifically, we have made a series of ex situ attempts to mimic the in situ natural spinning process in silkworm glands by adjusting the pH, changing the metallic ion concentrations, or applying shear force. The experimental results show that only the RSF nanofibrils created by shear forces adopt the para- $\beta$ -sheet arrangement (not yet published). Different  $\beta$ -sheet arrangements might contribute to the distinct mechanical properties between silk fibers and hydrogels. This is because parallel- $\beta$ -sheets are more regularly patterned along the fibrous axis; consequently, they are stronger. In this regard, the application of shear force during  $\beta$ -sheet formation might be a promising strategy for the preparation of various non-fiber SF materials with enhanced mechanical properties.

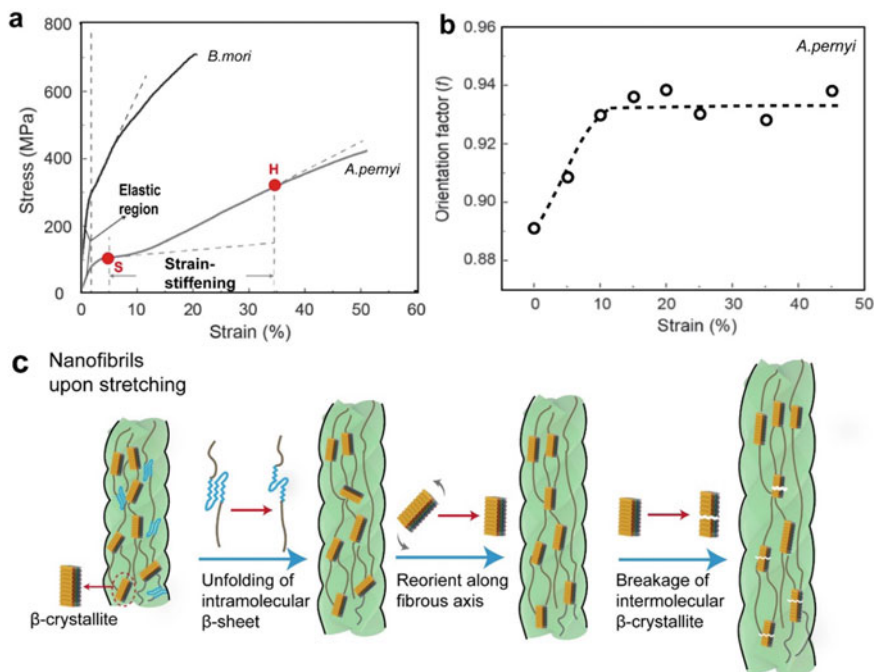
Although  $\beta$ -crystallites are mostly responsible for the outstanding mechanical performance of silk fibers, the intermolecular  $\beta$ -sheet is also equally important and plays a unique role. Specifically, mechanical tests show that *A. pernyi* silk fibers display a unique strain-stiffening characteristic [33, 43]. Recently, the structural origin of this strain-stiffening effect has been comprehensively studied [43]. The mechanism of strain stiffening is described as follows:

At relatively low stretching ratios (i.e., less than 5% or before the so-named yield point  $S$ , shown in Fig. 6.13a), the content of the intramolecular  $\beta$ -sheets in *B. mori* silk slightly decrease [33]. In comparison, the intramolecular  $\beta$ -sheet content in *A. pernyi* silk fibers decreases significantly [33, 43]. In other words, the intramolecular  $\beta$ -sheets in *A. pernyi* silk fibers are unfolded prior to the splitting of the intermolecular  $\beta$ -crystallites [33, 43]. This is attributed to the fact that the intramolecular  $\beta$ -sheets show a lower level of morphological perfection and stacking compactness than the intermolecular  $\beta$ -crystallites. Thus, the unfolding of intramolecular  $\beta$ -sheets results in the release of the entire length of the protein chains, which consequently leads to the extension of draglines without causing the breakage of the intermolecular linkage of molecular networks (i.e., the  $\beta$ -crystallites). In terms of macroscopic



◀**Fig. 6.12** Schema of crystal network of silk materials. **a** AFM morphology of regenerated nanofibrils in SF solutions. **b** Typical force-extension curves and schematic structure of amyloid fibrils. Reproduced with permission [35]. Copyright 2006, Elsevier. **c** Results reported by Oroudjev et al.; heights of peaks are random. Reproduced with permission [36]. Copyright 2002, National Academy of Science. **d** Latest report of force spectra and corresponding mechanism (**e**) Molecular fishnet structure of a silk nanofibril. **e** Simulated mechanical strength of fishnet, no friction strings, and entangled amyloid fibril structures of silk fibrils. Breaking forces are plotted as functions of length of fibrils, i.e., number of rows along fibril axes. Insets show a drawing of the different types of networks. **f** Microscopic mechanism of silk yielding behavior and effect of ordering function of crystallite network. (i) illustrates the interspersed  $\beta$ -crystallites in nanofibrils for which the orientations deviated from those of the fiber axis, (ii) illustrates splitting of  $\beta$ -crystallites in an arbitrary direction, and (iii) is a sketch presenting the number of H-bonds involved in splitting dynamics, indicated using red dashed lines. Better alignment along fiber axis would cause more H-bonds to be recruited to undergo the splitting force and consequently give rise to stronger crystalline binding interactions within  $\beta$ -crystallites. **d–f** Reproduced with permission [18]. Copyright 2016, Wiley-VCH

performance, during the above process (strain less than 10%), the modulus of the silk drops to nearly zero, as fiber extension is mainly caused by the breaking of weak intramolecular hydrogen bonds within the intramolecular  $\beta$ -sheets. Afterwards, with the progressive unfolding of fibers, the intermolecular  $\beta$ -crystallites begin to support the load. The silk fibers become stiffer owing to the contribution of the enthalpic component [33]. This is how strain-stiffening occurs. It was also found that the intermolecular  $\beta$ -crystallites in *A. pernyi* silks are reoriented and become aligned along the fibrous axis (Fig. 6.13b) [43], which further contributes to the stiffening of the entire crystal network (Fig. 6.13c). The stretching of fibers beyond inflection point H causes a failure in the function of  $\beta$ -crystallites. Demolishing nodes of the crystal network in fibers results in the weakening of the fiber (called strain weakening). This breakage process has been verified by XRD measurements, which showed that once the silk is stretched beyond point H, the  $\beta$ -crystallinity drops immediately [33]. *B. mori* fibers with significantly lower intramolecular  $\beta$ -sheet content in non-crystalline regions (only 9%) have lower flexibility and exhibit only strain-weakening behavior after the yield point. Thus, it is possible that if there is a method for increasing the number of intramolecular  $\beta$ -sheets in SF materials, the mechanical properties should be reinforced accordingly. However, it should be noted that the content of intramolecular  $\beta$ -sheets is greatly controlled by intrinsic primary structures. As discussed above, the core repetitive motifs forming  $\beta$ -sheet crystallites in *B. mori* and *A. pernyi* silks are  $(\text{GAGAGS})_n$  and poly-alanine motifs, respectively. In addition, it was confirmed that the  $(\text{GAGAGS})_n$  motifs in *B. mori* silk fibers are significantly longer. Longer lengths of repetitive  $(\text{GAGAGS})_n$  sequences can increase the likelihood of their association with each other and consequently lead to the formation of intermolecular  $\beta$ -crystallites instead of intramolecular  $\beta$ -crystallites. On the other hand, a statistical comparison of the primary fibroin sequences of *B. mori* and *A. pernyi* further revealed that the *B. mori* silk fibroin contains approximately 65% of the repetitive GAGAGS motif, whereas *A. pernyi* contains only approximately 18% of the repetitive poly(A)



**Fig. 6.13** **a** Typical stress-strain curves of *A. pernyi* and *B. mori* silkworm silk fibers. The points S and H are denoted as the yield and inflection points, respectively. The strain stiffening effect was only observed in *A. pernyi* silk fibers. **b** Variations in orientation functions of *A. pernyi* silk as a function of strain. **a, b** Reproduced with permission [43]. Copyright 2017, Wiley-VCH. **c** Schematic illustration of response of stretched *A. pernyi* silk fibers. It consists of three distinct stages: (i) unfolding of intramolecular  $\beta$ -sheets, which makes the fiber extensible (this stage is not observed in *B. mori* silk), (ii) reorientation of  $\beta$ -crystallites, which enhances the strength of fibers, and (iii) breakage of  $\beta$ -crystallites, which results in breakage of entire fiber

motif. This difference is attributable to the distinct content and size of the intramolecular  $\beta$ -crystallites in *B. mori* and *MA* silk fibers. Within the framework of the primary structure, the protein sequences might need to be altered to achieve an increment in the intramolecular  $\beta$ -sheet content of *B. mori* silk fibers and SF materials.

The orientation of crystallites is another important structural factor that determines the mechanical properties of crystal networks. However, although most of the  $\beta$ -strands in crystallites are roughly parallel to the fibril axis, the  $\beta$ -crystallites in the crystal network are still not perfectly oriented. So far, the influence of crystallite ordering on the breaking stress of silkworm silk has been intensively investigated via modeling and simulations, as well as experimental studies [18, 32]. According to the  $\beta$ -crystallite splitting theory, a better orientation of  $\beta$ -crystallites along the fiber axis results in higher breaking stress because it influences the effect force in the separation of  $\beta$ -sheets, as schematically illustrated in Fig. 6.12g [18].

In terms of experimental results, it was revealed that the mechanical performance of *B. mori* silk fibers can be improved by increasing the orientation function value. For example, Liu et al. [18] carried out MD simulations to investigate the influence of different orientation functions on the stability of corresponding crystal networks and then compared the simulation results with the experimental results. Notably, the simulated data fit well with the experimental data; that is, the breaking stress of *B. mori* silk fibers increases with  $f$  [18].

Apart from the crystallite orientation function, the density of crystallites is another important structural factor that determines the strength of silk fibers. Evidently, within the framework of the crystal network structure, the fishnet-like topology of this crystal network can further isolate the products of the breakage of individual crystallites by bypassing the loading stress from the broken ones to the surrounding interconnected crystallites [18]. Consequently, the more  $\beta$ -crystallites that participate in load sharing, the stronger the network.

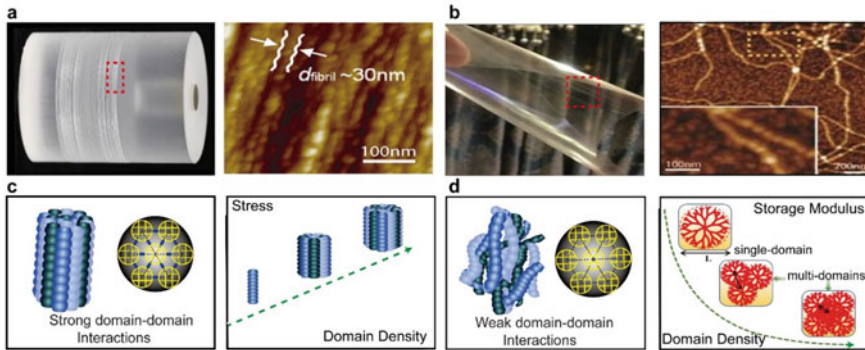
### 6.3.5 Level Five Structure of SF Materials

As discussed in Sect. 3.4, each individual nanofibril is indeed a molecular crystal network. If a nanofibril is isolated, we can treat such nanofibrils as a single, individual domain. If several nanofibrils are interconnected with each other and consequently form a nanofibril network, they are treated as a multi-domain system [19, 32]. In principle, the nanofibril network is defined as the level 5 structure of SF materials in which the structural unit refers to the individual nanofibril, and the links between neighboring nanofibrils are formed either by strong physical contact (for silk fibers) or weak interactions (for non-fiber SF materials). It is apparent that the stability of the nanofibrils themselves and the strength of inter-nanofibril interactions are the two most important factors that are highly correlated to the macroscopic mechanical properties [16, 19, 32]. In Sect. 6.3.4, the structural factors affecting the stability of nanofibrils (crystal network) are intensively investigated. In this subsection, we will comprehensively study the influence of the strength of interactions between nanofibrils.

Figure 6.14 illustrates two typical nanofibril network architectures in SF materials: (a) systems in which the inter-nanofibril interaction strength is strong or infinite, and (b) systems in which the inter-nanofibril interaction strength is weak or zero. AFM morphological results have shown that although both silk fibers and SF hydrogels are composed of numerous nanofibrils, the inter-nanofibril interaction strength in SF hydrogels is very weak because the nanofibrils in SF hydrogels are patterned in a disorderly manner. However, SF fibers exhibit a system of strong domain-domain interactions because SF fibers comprise a bundle of very well-oriented twisted nanofibrils interlocked by adjacent nanofibrils to ensure that the nanofibrils cannot move freely [16, 19, 32].

The above discussion reveals the importance of the helically twisted morphology of SF nanofibrils. We have assumed that this characteristic morphology results from

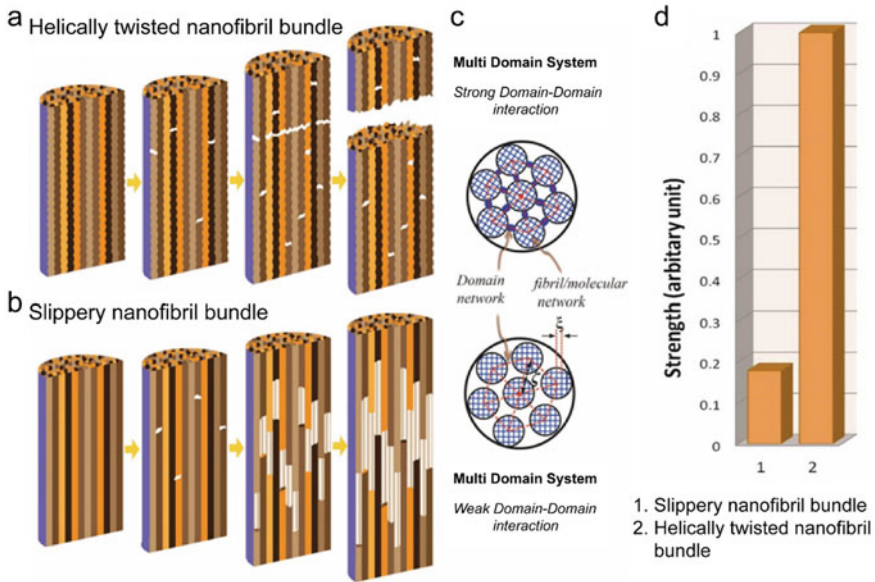




**Fig. 6.14** Illustration of level 5 structure of silk materials. AFM images reveal that SF nanofibrils are helically twisted **a** in silk fibers and **b** non-fiber silk materials such as SF films. **c** and **d** Illustration of the relationship between mechanical properties and density of nanofibril domains. **c** Weak inter-nanofibril interactions. **d** Strong inter-nanofibril interactions. Reproduced with permission [37]. Copyright 2019, Wiley-VCH

the associations between periodically repeating segments. Hence, the breakage of an individual nanofibril should take place at the most loaded segments. The fracture of entire SF fibers, which are composed of numerous nanofibrils, is initiated because of failure in the function of the weakest nanofibrils. However, the rough morphology of helically twisted nanofibrils can enable them to effectively avoid mutual slips, even after a critical external force has been applied. This is attributed to the presence of a non-slipping fibril bundle (N-SFB) structure; in addition, strong interactions between nanofibrils consequently give rise to stronger silk fibers.

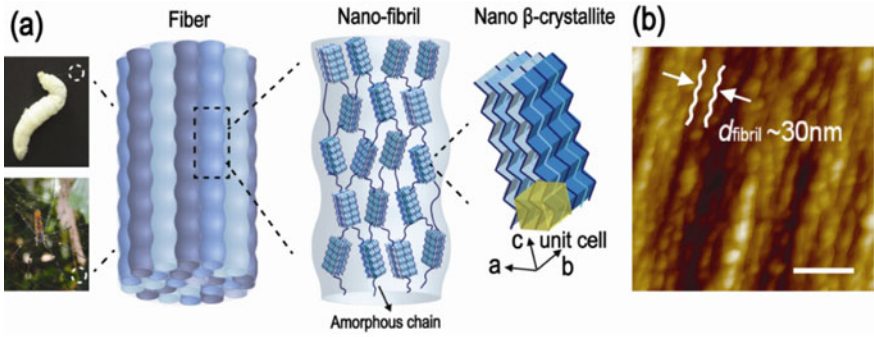
The advantage of this “non-slipperiness” feature in nanofibrils when toughening silk fibers can be demonstrated by comparing fibers with two other structures: the slippery fibril bundle (SFB) structure and the bulk network (BN) structure. The SFB model assumes that the nanofibrils are smooth and can slip out freely. The BN model assumes that the silk fibers are entirely bulk molecular crystallite networks. Using simulations, the mechanism by which the SFB and N-SFB structures react upon being stretched until breakage and the respective estimated mechanical strength were identified, as demonstrated in Fig. 6.15a and b. The manner of breakage of the BN structure is similar to that of brittle materials, and the emergence of fracture of the entire fiber results from the catastrophic growth of numerous small cracks. According to Griffith’s fracture theory, the extra stress caused by a small crack would mainly cause it to dissipate adjacently, especially at the transverse boundary [32]. In this regard, within the framework of the BN model, the extra stress is redistributed uniformly in the cross-sections that contain cracks. The accumulation of such extra stress can promote the formation of cracks, which finally results in the splitting of the entire network. In comparison, in the SFB structure model, a bundle of smooth nanofibrils is stretched gradually. As slipperiness results in weak interactions between neighboring nanofibrils, each nanofibril carries a certain load independently; thus, the breakage of the nanofibril bundle (the entire fiber) begins from the weakest



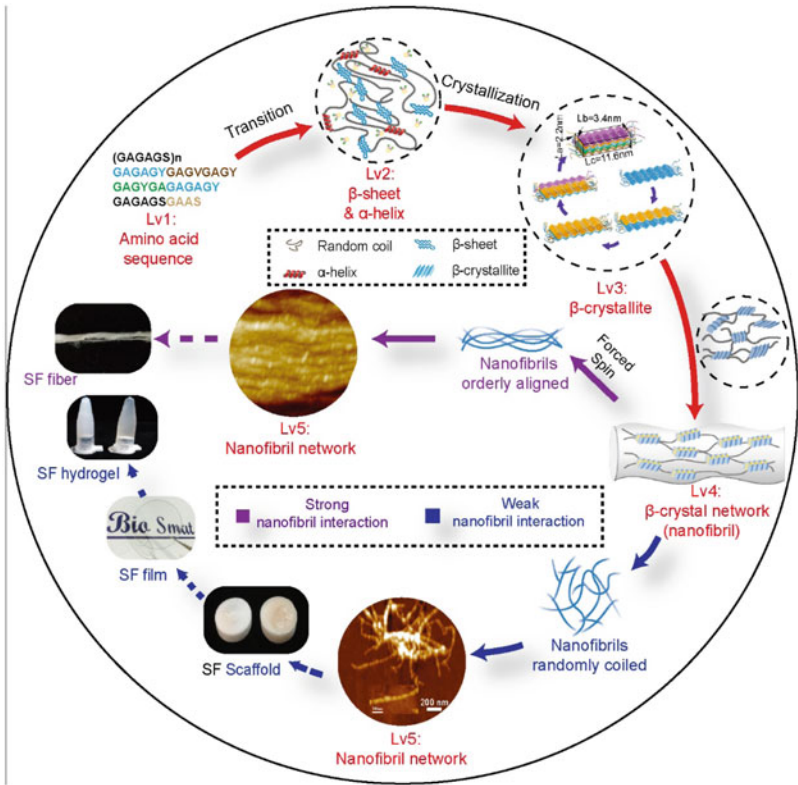
**Fig. 6.15** Correlation of nanofibril bundles with the breakage of silk fibers. How SF nanofibrils correlate will determine breaking pathway and toughness of silk fibers. Schematic illustration of breakage of silk fibers with **a** helicly twisted nanofibril bundle structure and **b** slippery nanofibril bundle structure. **c** Periodic segmental morphology of helicly twisted nanofibrils gives rise to friction and consequently to strong domain-domain interactions. In comparison, relatively weaker domain-domain interactions are observed in the slippery nanofibril bundle model. **d** Comparison of breaking strength of silk fibers, simulated according to above two models. Reproduced with permission [32]. Copyright 2014, Royal Society of Chemistry

nanofibril. After the weakest nanofibril breaks, the total loads are passed onto the adjacent nanofibrils, which eventually gives rise to the catastrophe of the unbroken nanofibrils, leading to the breakage of the entire fiber. The abovementioned process is similar to that seen in ductile materials, in which deformation is carried out by localized shear forces at the nanoscale level. The simulated breaking strengths of silk fibers are plotted in Fig. 6.15d. The SFB fibers are much weaker than those in the N-SFB model. In the BN model, a crack can easily develop at the weakest position along the transverse direction. However, in the N-SFB model, the boundaries of the helicly twisted nanofibrils can physically terminate the growth of such cracks across fibers, and the extra stress can be uniformly redistributed among surviving nanofibrils. However, in the SFB model, owing to the lack of strong friction between neighboring nanofibrils, cracks can easily occur along the fibrous axis without any impediments. In summary, the helicly twisted morphology of nanofibrils results in strong inter-nanofibril interactions, which further stops the occurrence of cracks in the transverse and longitudinal directions.

Based on the above discussion, there are at least two impressive structural factors in the level 5 structures of silk materials: (1) the periodic segmental morphology of



(A)



(B)

◀**Fig. 6.16** **A** Hierarchical Structures of spider and cocoon silk fibers; Reproduced with permission [32]. Copyright 2014, the Royal Society of Chemistry. **B** Schema of hierarchical network structures of SF fibers and non-fiber SF materials. (i) Level 1: amino acid sequence; (ii) Level 2:  $\alpha$ -helix &  $\beta$ -sheet. They are stabilized by intermolecular H-bonds, and  $\beta$ -sheets are crystallized. (iii) Level 3:  $\beta$ -crystallites. The formation is attributed to inter-sheet interactions, and several neighboring  $\beta$ -sheets (from different molecules) can crystallize into intermolecular  $\beta$ -crystallites (iv) Level 4: crystal network. It is composed of numerous  $\beta$ -crystallites, which are connected to each other by amorphous chains. A crystal network is indeed an individual nanofibril. During the process of crystal network formation, shear forces can help to orient directions of crystallites. (v) Level 5: nanofibril network. Based on nanofibril network topology and inter-nanofibril interaction strength, silk fibers and non-fiber silk materials are of different fibril arrangements: among silk fibers, silk nanofibrils are bundled along the fibrous axis, while for non-fiber silk materials, silk nanofibrils are interconnected in a nearly random manner. Reproduced with permission [37]. Copyright 2019, Wiley-VCH

nanofibrils, which prevents adjacent nanofibrils from slipping, and (2) the nanofibril bundle architecture as well as the non-slipperiness between nanofibrils causes extra stress to be shared equally among the unbroken nanofibrils.

As shown in Fig. 6.14, nanofibrils in silk fibers and SF hydrogels have similar diameters. In addition, most structural parameters, including level 1 to level 4 structures, except the orientation of  $\beta$ -crystallites, are also nearly the same [18]. However, different types of interactions between nanofibrils are observed in these two level 5 SF materials. Specifically, weak nanofibril–nanofibril interactions are observed in non-fiber SF materials, while strong inter-nanofibril interactions are observed in silk fibers. They both exhibit different, distinct nanofibril network architectures. For instance, the nanofibrils in silk fibers are well aligned along the fibrous axis, whereas the nanofibrils in regenerated SF hydrogels are much less ordered or randomly distributed. Hence, natural silk fibers are much stronger than SF films (measured at approximately 300–400 MPa and 40 MPa for silk fiber and SF films, respectively).

### 6.3.6 Summary of Hierarchical Structure of SF Materials

Based on the structural analyses of SF materials, the extraordinary properties of *B. mori* silkworm silk materials are attributed to the five levels of hierarchical network structures illustrated in Fig. 6.16. Figure 6.16a shows the key hierarchical structures of cocoon silk fibers. In detail and following a bottom-to-top manner, the hierarchical network structure can be summarized as follows: (1) amino acid sequence, (2) secondary structure, (3)  $\beta$ -crystallites, (4) crystal networks or nanofibrils, and (5) nanofibril networks.

Within the framework of the hierarchical structure, the crystalline binding interaction is very important in the stabilization of SF materials. The crystallite networks or nanofibrils and the networks of nanofibrils (Fig. 6.16b) are two of the most essential structural elements, which, to a large extent, determine the macroscopic performance of SF materials.

## 6.4 Characterization Technologies

As discussed in Sect. 6.3, an accurate and quantitative method for the characterization of the secondary and tertiary structure content within SF materials is of utmost importance. In addition, to provide comprehensive insight into the morphology and architecture of the nanofibrils as well as the nanofibril network, advanced imaging techniques are also in demand. So far, substantial effort has been devoted to the development of corresponding techniques, and several structural characterization and imaging methods have been developed. For instance, structural characterization methods include FTIR spectroscopy [47, 48], Raman spectroscopy [49, 50], circular dichroism (CD) [51], and XRD spectroscopy [33]. In addition, AFM force spectra have also emerged as powerful tools [18, 52] for studying the nanostructures within SF materials. Scanning electron microscopy (SEM), transmission electron microscopy (TEM), and AFM are the most popular imaging techniques. In this section, we provide a brief introduction to the principles and operation of the aforementioned methods; specifically, several typical examples are also presented.

### 6.4.1 Structural Characterization Techniques

#### 6.4.1.1 Overview of Structural Characterization Techniques

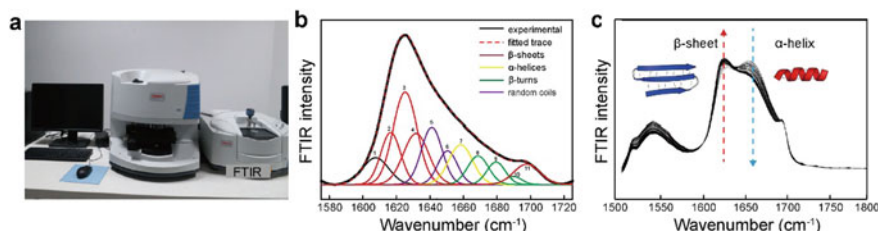
As introduced in Sect. 6.3, the total  $\beta$ -conformation can be classified into intramolecular  $\beta$ -sheets ( $\beta$ -crystallites) and intermolecular  $\beta$ -sheets, respectively. Given the essential role of both  $\beta$ -crystallites and  $\beta$ -sheets, precise measurements of their contents are of the utmost importance. In general, FTIR, Raman, and CD spectroscopy are used to quantify the total secondary structure content, while XRD is the most commonly applied technique for analyzing the level 3 structure of SF materials [8, 53]. Moreover, polarized Raman spectroscopy is also capable of providing information on the molecular orientation of SF materials. Wide-angle X-ray scattering (WAXS), on the other hand, was applied to calculate the  $\beta$ -crystallite size, crystal density, and crystal orientation. SAXS is powerful for measuring the distance between adjacent  $\beta$ -crystallites in both the horizontal and vertical directions.

#### 6.4.1.2 Fourier Transform Infrared Spectrometer (FTIR)

Owing to its simplicity and practical usage, FTIR is the earliest and one of the most widely applied methods for studying the secondary structures of proteins. FTIR spectroscopy can be used to obtain information from the infrared spectrum of molecular vibrations and rotational information. Its principle is that when the bonds between the atoms in protein polypeptides (e.g., C=O bonds and C–N bonds) stretch and

bend, they will absorb the infrared energy and consequently display a characteristic spectrum. Because the vibration frequencies for different bonds are distinct, the corresponding spectrum can therefore be regarded as a fingerprint of the molecules that can be used for identification. Peak deconvolution analysis is usually required for SF material characterization using FTIR to quantify the content of each secondary structure. Specifically, deconvolution is carried out on the amide I vibration band, which is composed of strongly overlapping components that correspond to various secondary structures, including the  $\beta$ -sheet,  $\alpha$ -helix,  $\beta$ -turn, and random coil. The amide I vibration band is chosen because it is the only band among all amide vibration bands that depends on the secondary structure of the protein backbone; thus, it is minimally affected by the nature of the side-chain residues. Using peak deconvolution methods, the amide I band in the FTIR spectra can be fitted with several Gaussian peaks indicative of different secondary structures. For example, the peaks corresponding to the  $\beta$ -sheet are located at approximately  $1620$  and  $1698$   $\text{cm}^{-1}$ , whereas the amorphous components are located at approximately  $1645$   $\text{cm}^{-1}$ . The content of each secondary structural component can be determined by measuring the ratios of the areas under the corresponding peaks in those areas (Fig. 6.17b) [47].

Previously, most of the studies using FTIR for SF material structural characterization have yielded static instead of dynamic data, as they only focus on the conformations of SF molecules before and after the natural spinning/gelation process, or under other specified conditions. However, it is useful to monitor the conformation transition kinetics continuously in time [47, 48]. Recently, time-resolved FTIR spectroscopy has been promoted for monitoring the kinetics of conformational transitions induced by various environmental factors (Fig. 6.17c). For instance, the time-resolved FTIR process has been applied to study the influence of metallic ions on the conformation transition process in dried spidroin/fibroin films [47]. Similarly, the conformation transition kinetics of SF films and SF aqueous solutions induced by changes in pH and organic solvents with a low dielectric constant (e.g., methanol and ethanol) have also been studied using time-resolved FTIR [48].



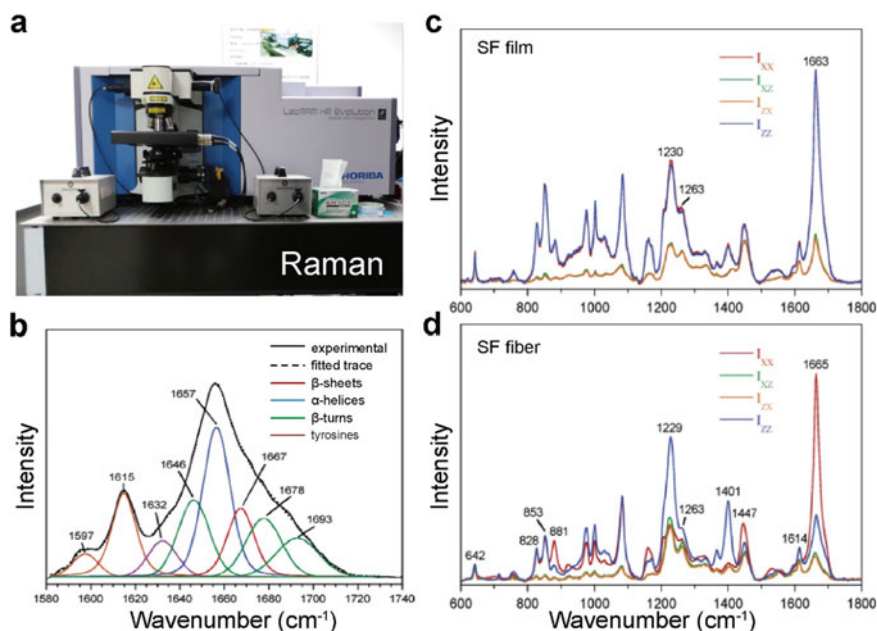
**Fig. 6.17** FTIR technique for characterizing secondary structure in silk materials. **a** Image of FTIR instrument. **b** Deconvolution of FTIR spectra and assignment of peaks. **c** Conformation transition kinetics of regenerated *B. mori* silk fibroin membrane monitored by time resolved FTIR spectroscopy. Reproduced with permission [47]. Copyright 2001, Elsevier

### 6.4.1.3 Raman Spectroscopy

Other than FTIR, Raman spectroscopy is another powerful and nondestructive technique used to investigate the secondary structures of SF materials [49, 50]. In general, Raman spectroscopy can provide information that is complementary to that of FTIR spectroscopy. As discussed, the signals in the FTIR spectrum correspond to the absorbance energy of the infrared photons released by molecules. In contrast, the Raman effect deals with the scattering process involving interactions between the incident photons and the sample molecules; in other words, the Raman spectra record the inelastically scattered energy of photons. For molecules to be infrared-active, molecular vibrations must initiate a change in the dipole moment of the molecules. To determine the Raman activity, the molecular vibrations should induce a change in the molecular polarizability. In principle, if a molecule has a center of symmetry, then a Raman-active vibration is always infrared-inactive, and vice versa [54]. For example, polarizable bonds such as C–C, S–S, N=N, and O–O bonds can display intense Raman bands; however, in the infrared spectrum, these bonds show only weak or even undetectable bands. Another advantage of Raman spectroscopy for studying protein molecular structure is that it can also provide quantitative information about the content of secondary structures in a manner similar to the FTIR method (Fig. 6.18b).

Recently, a Raman spectromicroscopy technique was developed by mounting a microscope onto the conventional Raman spectroscopy setup. This technique can precisely refine the size of an incident laser beam to less than 20  $\mu\text{m}$  and can thus collect scattered signals from very small samples, making it an ideal tool for recording the high-quality Raman spectra of single silk filaments. Moreover, by combining Raman spectromicroscopy with an external mechanical deformation puller, the transition process of secondary structures in silk filaments when subjected to mechanical stretching can be detected in situ. Specifically, it has been reported that silkworm silks display well-defined Raman spectra in which the frequencies of some bands shift under the action of tensile stress or strain, suggesting a molecular conformation transition. For instance, Raman microspectroscopy has been used to quantitatively study the effect of mechanical deformation on the secondary structure conformation and order parameters of *Samia cynthia ricini* (*S. c. ricini*) silk fibroin fibers. According to a study by Rousseau et al. [55] samples were obtained from an aqueous solution stored in the silk gland and stretched at draw ratios ( $\lambda$ ) ranging from 0 to 11. The Raman data unambiguously show that in response to mechanical deformation, SF molecules undergo a cooperative  $\alpha$ -helix to  $\beta$ -sheet conformational transition.

Based on the FTIR and Raman data, it is surprising that the amounts of  $\beta$ -sheets in silkworm silk fibers (~50%) nearly coincide with the proportion of relevant amino acid sequences (i.e., GAGAGS amino acids that are recognized to be involved in  $\beta$ -sheet generation), which is 53%. In comparison, for *Nephila* dragline silk fibers, the amino acids (A)<sub>n</sub> constitute only 18% of the total content, and this is significantly lower than the  $\beta$ -sheet content (36–37%) measured by Raman spectroscopy. In the AG and GGA motifs, which are usually located adjacent to the (A)<sub>n</sub> blocks, the sum value increases to 31%. Thus, this result strongly suggests that apart from the poly

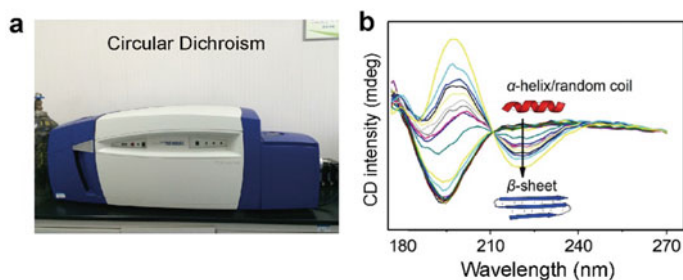


**Fig. 6.18** Raman spectroscopy for characterizing secondary structure and polarized Raman spectra to measure orientation in silk materials. **a** Image of Raman spectroscopy. **b** Raman band decomposition of amide I region ( $1580\text{--}1740\text{ cm}^{-1}$ ) of isotropic spectra for *Samia cynthia ricini* (*S. c. ricini*) silk fibers. Reproduced with permission [55]. Copyright 2006, American Chemical Society. Polarized Raman spectra of the coagulated film and **c** The  $I_{xx}$  and  $I_{zz}$  curve are nearly the same. **d** Polarized spectra of the fibroin fiber of *B. mori* silk. The  $I_{xx}$  and  $I_{zz}$  curve are significantly distinct. **c, d** Reproduced with permission [50]. Copyright 2004, American Chemical Society

alanine sequences, the AG and GGA sequence motifs might also be incorporated into the  $\beta$ -sheets [56].

In addition to the molecular conformation, polarized Raman spectra have recently been adopted to compare the molecular orientation within SF materials [57]. In a typical polarized Raman spectroscopy measurement, the intensity  $I_{ij}$  refers to the intensity measured when the incident beam is polarized in the  $i$  direction and the scattered light is polarized in the  $j$  direction, where  $i, j = x, y, \text{ or } z$  are the axes associated with the laboratory frame. The polarized spectra in the amide I region of both the regenerated SF film and cocoon silk fiber are presented in Fig. 6.18c and d, respectively [57]. For the SF film, the intensity of the amide I band in the  $xx$  and  $zz$  spectra is equal, implying that the sample is isotropic. However, the intensity of the amide I band is much weaker in the  $zz$  spectrum of the cocoon silk compared to that in the  $xx$  spectrum, which is consistent with the high orientation of the peptide groups. Specifically, this is attributed to the fact that the carbonyl groups within the SF and SF films are mostly oriented along the  $x$  axis and are randomly distributed. In this way, by measuring the ratio of the intensity of  $I_{xx}$  and  $I_{zz}$ , the molecular orientation can be quantitatively measured [57].





**Fig. 6.19** **a** Image of CD spectroscopy **b** Conformational transition process of a RSF solution revealed by CD spectra

#### 6.4.1.4 Circular Dichroism Spectroscopy

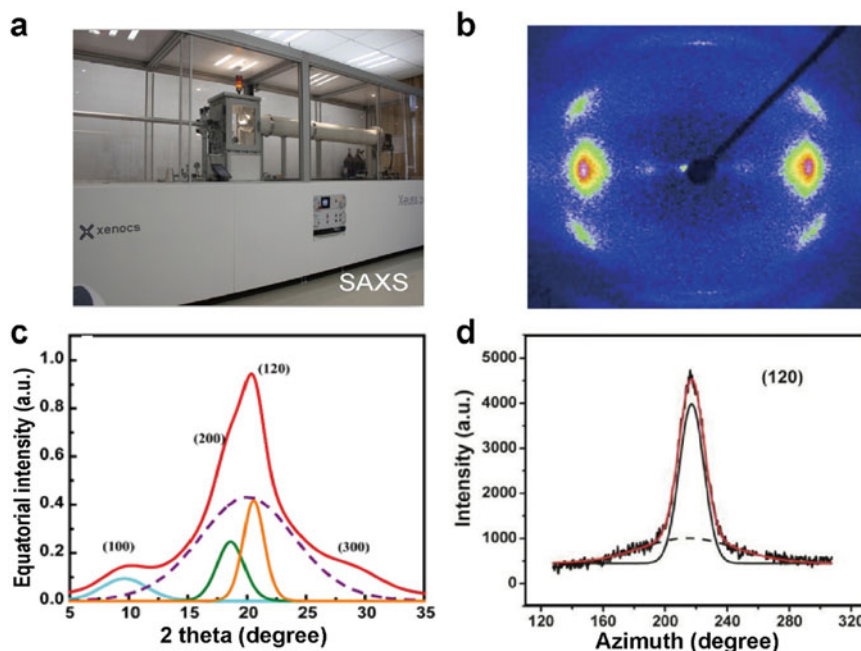
Compared to FTIR and Raman spectroscopy techniques, CD spectroscopy is more suitable for quantifying the secondary structural content of proteins dissolved in solutions [51]. CD spectroscopy is uniquely sensitive to protein chirality or overall asymmetry. It can measure the differential absorption between left-handed and right-handed circularly polarized light as a function of wavelength [58]. As the chirality of protein molecules is determined solely by the secondary conformation, information about the secondary structure can be obtained from CD measurements. Moreover, CD spectroscopy is also capable of monitoring the conformational transition kinetics in situ under different conditions (Fig. 6.19b). For instance, Canneti et al. [59] applied CD spectroscopy to investigate the conformational transition of fibroins in both aqueous solutions and in organic solvents. Dicko et al. [58] studied the influence of storage time, storage temperature, and methanol solvents on the conformation transition kinetics of spidroin. Li et al. [60] used CD spectroscopy to measure conformational transition kinetics and present evidence indicating that such transitions should follow a nucleation-dependent aggregation mechanism. It should be noted that although CD spectroscopy can serve as a versatile method for examining the secondary structure of SF solutions in principle, its accuracy is mostly affected by protein concentration. In practice, only very diluted SF solutions (i.e., with concentrations ranging from 0.01 to 0.2 g/L) are appropriate for CD experiments.

#### 6.4.1.5 Wide Angle X-Ray Diffraction (WAXD) and Small Angle X-Ray Scattering (SAXS)

The principle of XRD is that the crystalline structure within silk materials can cause incident X-ray beams to diffract into many specific directions. By measuring the angles and intensities of such diffracted beams, a two-dimensional profile reflecting the crystalline structures can be produced and gathered for quantitative analysis. In particular, according to Bragg's Law, in the case of sub-nanometer-sized crystalline structures (e.g., 0.1–10 nm), WAXD, and wide-angle X-ray scattering (WAXS) can

be applied. SAXS is ideal for obtaining the structural information of larger crystalline molecules with dimensions between 1 and 100 nm and of repeat distances in partially ordered systems with a length of up to 150 nm. Because silk materials belong to semi-crystalline macromolecular materials, WAXS has been developed as a powerful method for studying the size and orientation of crystallites as well as the crystalline content (crystallinity) within the crystal network of silk materials. The SAXS technique was applied to measure the inter-crystallite distances within the crystal network inside silkworm silk and spider dragline silk fibers [8]. The detailed analysis methods are described below.

*Crystallinity (crystallite content) and crystallite size:* Radial integration along the equatorial and meridian directions of the WAXS pattern (Fig. 6.20b) can form a 1D profile of the scattering intensity as a function of the  $2\theta$  angle (Fig. 6.20c). This profile is then deconvoluted into several peaks (corresponding to crystalline domains) and a halo (corresponding to amorphous regions). For silkworm silk fibers, the equatorial data are deconvoluted into four crystalline peaks corresponding to the (100), (200),



**Fig. 6.20** XRD analysis of the structural information of silk materials. **a** Image of SAXS. **b** Typical scattering pattern of silkworm silk fiber. **c** The radial integration of intensity as a function of diffraction angle ( $2\theta$ ) along the equatorial direction of the WAXS pattern. **d** Intensity as a function of azimuth angle at radial position of equatorial (120) peak of silkworm silk. The peak is fitted as sums of two Gaussians corresponding to crystalline (black, solid line) and amorphous (black, dashed line) distributions. The orientation function  $f$  is calculated by FWHM of the crystalline Gaussian peak

(120), and (300) Bragg reflections and an amorphous halo. The meridian data, on the other hand, is deconvoluted into two more crystalline peaks corresponding to (002) and (102) as well as an amorphous halo. The crystallinity is determined by the ratio of the area under the crystalline peaks in the equatorial data (i.e., the (100), (200), (120), and (300) peaks) to that of the total reflection patterns. According to Scherer's formula, the crystalline size in one dimension is equal to  $\frac{0.9\lambda}{FWHM\cos\theta}$ , where FWHM represents the full width at half-maximum of the peak at the diffraction angle  $\theta$ , and the wavelength of the incident ray  $\lambda$  is 0.15418 nm. The crystallite sizes along the  $a$ ,  $b$ , and  $c$  directions are determined by the position and the full width at half maximum (FWHM) of the (200), (120), and (002) peaks, respectively [8]. For spider dragline silk fibers, the crystallite size is measured to be  $a = 2.1$  nm,  $b = 2.7$  nm, and  $c = 6.5$  nm. The dimensions of silkworm silk fibers are relatively larger, i.e.  $a = 2.3$  nm,  $b = 4.1$  nm, and  $c = 10.3$  nm.

*Orientation Function  $f$* : Experimentally, information about the orientation of crystallites can be measured via the WAXS intensity integration as a function of the azimuth angle at the radial position of the equatorial (120) and (200) peaks (Fig. 6.20c). Here, the orientation function  $f$  is defined by the Hermans orientation function, as Eq. 6.1,

$$f = (3 \langle \cos^2 \phi \rangle - 1)/2 \quad (6.1)$$

where  $\phi$  is the angle between the  $c$ -axis of the crystallites and the fiber axis. For the two reflections, (200) and (120), which are not orthogonal but have a known geometry in the equatorial plane, the expression of  $\langle \cos^2 \phi \rangle$  is determined using the Eq. 6.2

$$\langle \cos^2 \phi \rangle = 1 - 0.8 \langle \cos^2 \phi_{200} \rangle - 1.2 \langle \cos^2 \phi_{120} \rangle \quad (6.2)$$

The FWHM values of the (200) and (120) peaks were measured in a direction perpendicular to that of the fiber axis using the following Eq. 6.3,

$$\langle \cos^2 \phi_{200} \rangle = 1 - [\cos(0.4FWHM_{200})]^2 \quad (6.3)$$

Thus, the FWHM data can be applied to calculate the orientation function. If  $f = 1$ , the  $\beta$ -crystallites are oriented in a direction that is completely parallel to the fiber axis. However, if  $f = 0$ , the  $\beta$ -crystallites are oriented randomly [8].

*Inter-crystalline distance*: As described by the molecular crystal network structure, crystallites, which scatter the incident X-rays, are embedded randomly into amorphous regions and have a cylindrical symmetry along the fiber axis. In this regard, the SAXS intensity in the equatorial direction can be determined using the Eq. 6.4

$$I = \frac{Kl_c^2}{(1 + l_c^2 q^2)^{3/2}} \quad (6.4)$$

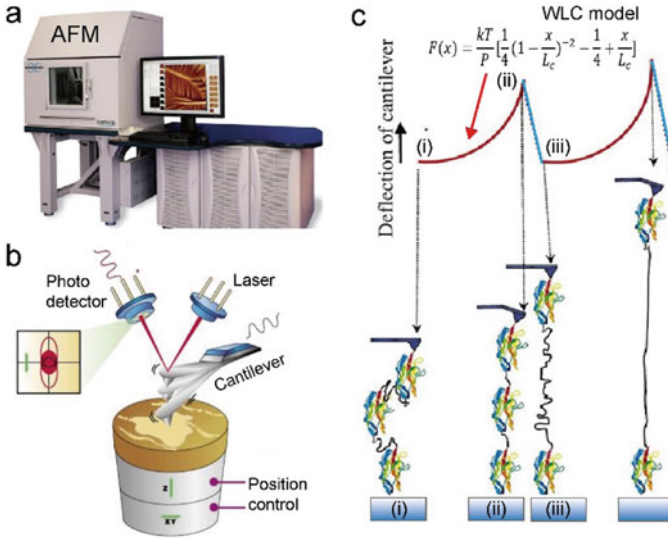
Here,  $q = 4\pi \sin\theta/\lambda$  is the scattering wave vector,  $\theta$  is the scattering angle,  $\lambda$  is the X-ray wavelength, and  $l_c$  is the correlation length, which can be determined from the intercept and slope of a plot of  $I^{-2/3}$  versus  $q^2$ . To determine the level of scattering along the meridional direction, the SAXS intensity is expressed as Eq. 6.5,

$$I = \frac{Kl_c^2}{1 + l_c^2 q^2} \quad (6.5)$$

Here, the correlation length  $l_c$  along the meridian is determined from the intercept and slope of the plot of  $I^{-1}$  versus  $q^2$ . These two correlation lengths can be related to the inter-crystalline distance. Notably, the silk fiber has a fibrous axis with different properties along the equatorial and meridional directions. Hence, the correlation lengths should be considered separately along the two directions. Recently, Du et al. [8] applied this SAXS technique to study the influence of different reeling speeds on the inter-crystallite distances within spider dragline silk fibers. According to their results, there exists a deflection point in the correlation between the inter-crystallite distance and the reeling speed. Specifically, when the reeling speed is below 10 mm/s, the distance between the crystallites (correlation length) increases with the reeling speed along both the meridional and equatorial directions. This is consistent with the fact that the fast reeling speed easily extends the relatively loosely packed amorphous chains between crystallites, which results in larger intercrystallite distances. Nevertheless, upon further increase of the reeling speed (i.e., >10 mm/s), the nanofibril segments start to merge together so that the distance between the crystallites becomes smaller. This merging phenomena of nanofibril segments has been observed with AFM, in which the observed particle size becomes larger while several hollow regions begin to appear.

#### 6.4.1.6 Atomic Force Microscopy (AFM) Force Spectroscopy

Conventionally, AFM force spectroscopy is a technique that endows probing the mechanical unfolding and refolding processes of proteins and DNA at the single molecule level. During the measurements, the AFM cantilever approached and snapped into the sample, and then retracted from the surface. At the same time, the cantilever deflection (correlated to the force applied onto the sample) versus piezo movement (correlated to the separation between the AFM tip and protein sample) was recorded, which is ultimately converted to the *force versus extension* curves of protein in response to mechanical pulling and can provide detailed mechanical and nanostructural information about samples. For instance, AFM force curves can reveal various mechanical properties of the sample, including adhesion, stiffness (modulus), rupture force, and indentation depth (how much the AFM tip penetrates in the sample at a given load, which reflects the hardness of the sample). Recently, AFM force spectroscopy has also been applied to study the sequential unfolding of nano- $\beta$ -crystallites within SF materials (Fig. 6.21) [18]. The analysis focused on how



**Fig. 6.21** **a** Image of AFM. **b** Schematic illustration of the AFM setup. The force is measured by the deflection of the cantilever and the extension can be calculated from the position of the cantilever. **c** The unfolding of a protein domain (or crystallite in case of SF materials) by an external force. When stress is applied onto samples, the protein domains will begin to unravel. As the distance between substrate and cantilever increases [from states (i) to (ii)] the protein elongates and the reduction of its entropy generates a restoring force that bends the cantilever. When a domain unfolds [state (iii)], the contour length of the protein increases, returning the force on the cantilever to near zero. Further extension again results in force on the cantilever (state (i) again). The entropic elasticity of proteins can be described by the worm-like chain (WLC) model of polymer elasticity (inset). This equation predicts the entropic restoring force ( $F$ ) generated upon extension ( $x$ ) of a protein in terms of its persistence length ( $p$ ) and its contour length ( $L_c$ ). The saw-tooth pattern of peaks on the force-extension relationship corresponds to sequential unraveling of individual domains of a modular protein like the one shown here. The number of peaks correspond to the number of domains. (adapted from <http://www.bio.unipd.it/~bubacco/synuclein.html>)

these crystallites break (i.e.,  $\beta$ -sheets were pulled off or  $\beta$ -strands were unzipped) and can provide insight into how such crystallites are associated with each other and form crystal networks.

## 6.4.2 Imaging Techniques

### 6.4.2.1 Overview of Imaging Techniques

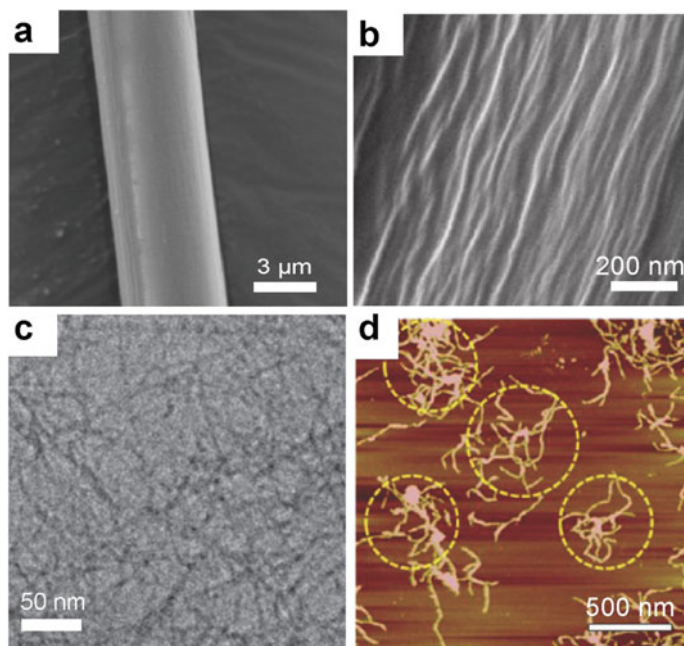
The structures of SF materials from levels one to three can be thoroughly investigated using the aforementioned structural characterization techniques. Because that nanofibrils and nanofibril networks are in the mesoscopic range, imaging techniques

can be applied to determine the morphology of nanofibrils in a straightforward and effective way. In general, SEM, TEM, and AFM are the three most popular imaging techniques. Although all of these techniques have been successfully applied for obtaining high-quality images of SF nanofibrils, they vary significantly in operation environment, horizontal/vertical resolution, and specific requirements for sample preparation. For instance, SEM and TEM must be operated in vacuum conditions. In comparison, AFM imaging can be applied in air, fluid, or vacuum conditions. SEM and TEM can provide information in the horizontal direction. However, AFM height imaging is capable of detecting the heights of samples at a very high resolution (0.1 nm). Because SF materials are non-conductive, they need to be coated with a layer of gold or platinum prior to SEM and TEM imaging.

#### 6.4.2.2 Scanning Electron Microscopy

SEM is capable of scanning SF material samples with a focused electron beam instead of light and delivering greatly magnified images with information about the topology of the samples as well as their composition. The principle of SEM is that the incident electrons can interact with the atoms within the SF material samples so that different parts of the samples produce various signals, that is, secondary electrons, backscattered electrons, and characteristic X-rays. These signals can reveal hidden information about the SF material samples. Specifically, the secondary electron signals are associated with the topography of the samples; the backscattered electrons can provide information about the phase contrast in the samples; and the characteristic X-rays can be applied for element identification (called the energy dispersive X-ray spectroscopy technique).

To date, SEM has been widely applied to examine the micro/nanostructures within different forms of SF materials. For instance, Nguyen et al. [34] imaged freeze-dried RSF hydrogels and found that they were composed of numerous nanofibrils with a random distribution. The corresponding SEM image clearly shows that the average diameter of the nanofibrils ranges from 20 to 50 nm [34]. For natural silk fibers, SEM has successfully distinguished the sericin coating from the core fibroin filaments (Fig. 6.22a) [6]. However, because silk fibers are non-conductive materials, they require a gold spraying process prior to SEM imaging. In addition, the nanofibrils within silk fibers are bundled very close to each other. For these reasons, it is much more difficult to obtain satisfactory images of the nanofibril network on the silk fiber surface. Nevertheless, upon proper sample preparation (e.g., freeze-drying and controlling spraying time), the morphology of nanofibrils within silk fibers can still be observed (Fig. 6.18b, not yet published).



**Fig. 6.22** Morphology of SF materials characterized by different imaging techniques. **a** Typical SEM image of natural silkworm silk fiber. **b** SEM image of natural silkworm silk fibers after proper freeze-drying treatment, revealing their fibrous morphology. **c** Cryo-TEM image of RSF gel (incubated 50 min) in which nanofibrils can be easily observed. Reproduced with permission [46]. Copyright 2009, Royal Society of Chemistry. **d** AFM height image of SF aqueous solution, in which nanofibrils have spontaneously formed and deposited on the substrate. Reproduced with permission [34]. Copyright 2015, Wiley-VCH

### 6.4.2.3 Transmission Electron Microscopy

Similar to SEM, TEM is a microscopy technique that also applies a beam of electrons, but it is capable of imaging at even higher resolutions. The electrons are transmitted through the samples to form an image. In principle, the samples for TEM measurements are most often ultra-thin sections less than 100 nm thick or suspensions on a grid, which makes it difficult to prepare ideal silk material samples. For instance, for cross sections of natural silk fibers, given the fact that  $\beta$ -crystallites within crystal networks are patterned parallel with each other and along the fibrous axis, the silk fiber displays a strong tenacity. In other words, when the silk fibers are perpendicularly sliced, the crystallites are more likely to be bent or deformed rather than torn apart. However, when applying the TEM technique to study the morphology of nanofibrils within SF gels, Gong et al. [46] have observed that the spontaneously formed translucent gel is composed of entangled proto-fibrils with lengths of hundreds of nanometers and a width of approximately 5 nm. In addition, they have also mimicked the flow effect by applying circular agitation to the

RSF solution and found that quickly several white fibrous flocs were generated after shearing the RSF solution, suggesting that shear flow has an impact on the formation of SF nanofibrils [46].

#### 6.4.2.4 Atomic Force Microscopic (AFM) Imaging

AFM can provide a nanoscale 3D profile of a sample surface by measuring the forces between a sharp probe (with a radius of less than 10 nm) and the sample surface. Different surface topographies and properties of samples can lead to different interaction values, which can be reflected and detected by the AFM instruments. To date, AFM height imaging techniques have been widely applied to study the morphology of silk materials. An advantage of AFM is that it can provide height information for the samples at a very high resolution. It is determined that the thickness of nanofibrils is much smaller than the diameters, suggesting that the SF nanofibrils are flat and ribbon-like rather than symmetrical cylinder-like aggregates.

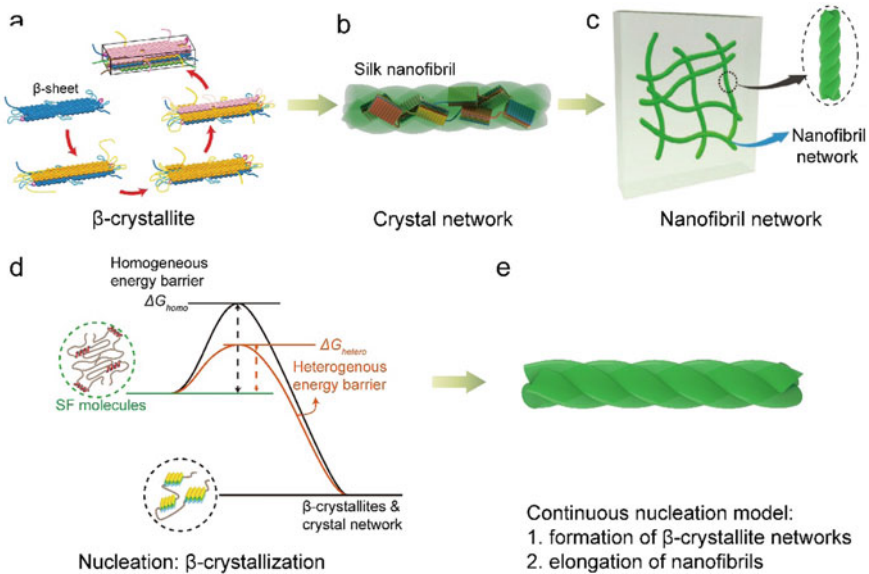
Recently, AFM has also been applied to distinguish the locations of different constituents within composite RSF films. For instance, Xing et al. [61] have incorporated wool keratin molecules @ gold nanoclusters (WK@AuNCs) into the mesoscopic network structures of SF films and synthesized novel bio-degradable WK@AuNCs-SF memristors and then successfully identified the location of such clusters via AFM. As the WK@AuNCs accumulate electrons on their surfaces, interactions with the conductive AFM tip can significantly differ from those of the SF molecules. In this regard, Kelvin probe force microscopy (KPFM) is capable of imaging potential differences by measuring the surface charge distribution [61].

## 6.5 Self-assembly Kinetic Pathways of Silk Materials

### 6.5.1 *Key Mesoscopic Structural Elements: Crystallites, Crystal Networks, and Nanofibril Networks*

From the introduction, it is clear that the mesoscopic hierarchical structure, to a large extent, determines the macroscopic mechanical properties of SF materials. In addition, the crystalline binding force turned out to be the most important factor in stabilizing SF materials, while the nano-fishnet topology crystal networks and nanofibril networks are another two of the other most essential structural elements at the mesoscopic scale (Fig. 6.23). Specifically, after the formation of large amounts of  $\beta$ -crystallites, a crystal network (also known as a nano-helical fibril) is spontaneously produced. Nevertheless, individual nano-helical fibrils are insufficient for maintaining the structural integrity of SF materials. Several nearby nano-helical fibrils can associate with each other and consequently form a nanofibril that displays a nanorope structure. Finally, the gathering of the aforementioned nanofibers together





**Fig. 6.23** Illustration of hierarchical mesoscopic structures of SF materials. The last three levels, which refer to the  $\beta$ -crystallites, crystal networks, and nanofibril networks, are key structural factors that mainly influence macroscopic mechanical performance of SF materials. **a** Schematic illustration of crystallization in controlling SF meso network formation. This is key principle in controlling mesoscopic structure engineering. Crystallization of  $\beta$ -sheets from different molecules gives rise to formation of intermolecular  $\beta$ -crystallites. **b** This leads to the formation of  $\beta$ -crystallite networks (nanofibrils). **c** Nanofibrils will further interact to bundle or entangle to form nanofibril networks, which play a key role in the macroscopic properties of SF materials. **d** Heterogeneous nucleation can lower energy barrier so as to accelerate the nucleation kinetics. **e** Continuous nucleation model: due to the fact that  $\beta$ -crystallites are very small, the nucleation becomes the rate limiting step in crystallization. Therefore, the formation of  $\beta$ -crystallite networks is resulted from the continuous nucleation of  $\beta$ -crystallites, along the direction of molecular chains. This gives rise to the elongation of nanofibrils

comprises the highest level of structure, that is, the nanofibril network. The formation of these mesoscopic structures is essential for endowing SF materials with outstanding mechanical properties as well as structural stability. For instance, silk fibers and SF films with small crystallinity can easily break or dissolve in water, respectively.

By studying the hierarchical network structures of SF materials and the corresponding formation kinetics, we have proposed another refined self-assembly pathway for SF molecules, which is, a continuous nucleation model. In this model, SF molecules self-assemble into hierarchical network structures via several nucleation-controlled steps. Specifically, these require (1) SF molecules to assemble and nucleate into  $\beta$ -sheets; (2) the  $\beta$ -sheets to grow into  $\beta$ -crystallites through layer-by-layer stacking; (3) the formation of  $\beta$ -crystallite networks is resulted from the continuous nucleation of  $\beta$ -crystallites; and (4) with continuous nucleation of  $\beta$ -crystallites along

the direction of molecular chains, nanofibrils elongate and give rise to interaction with each other and consequently lead to the formation of nanofibril networks. In detail, during stage 1, most SF molecules are in the random coil or  $\alpha$ -helix conformations and are therefore soluble. Gradually, owing to thermal fluctuations, such molecules begin to move forward towards each other. Along with the breakage of the initial intramolecular H-bonds and the formation of intermolecular H-bondings, the transition to  $\beta$ -sheets is triggered. The total free energy of the entire system is decreased because of the formation of  $\beta$ -sheets, and the  $\beta$ -sheets have a more compact structure than their precursors. In this regard, such a conformation transitional process can be identified as a crystallization process. During stage 2, because more  $\beta$ -sheets are present in the solution, inter-sheet interactions (such as hydrophobic interactions and van der Waals interactions) further lead to the stacking of nearby  $\beta$ -sheets and result in the crystallization of  $\beta$ -crystallites.  $\beta$ -crystallites are attributable to the closed package and well-defined patterns of  $\beta$ -sheets and should be regarded as more stable polymorphs of  $\beta$ -sheets. During stage 3, the shear force (such as the force in the silk spinneret) orients the  $\beta$ -crystallites in a parallel direction. However, in the absence of such shear forces (e.g., during the SF hydrogelation process), the orientation of crystallites is less orderly. In stage 4, the differences in nanofibril interaction strength and the nanofibril network architecture result in the differing mechanical performance of silk materials.

## 6.5.2 Nucleation Mechanism

The formation of the mesoscopic hierarchical crystalline network structure of SF materials has been found to follow the nucleation mechanism [19]. The nucleation kinetics of both the homogeneous nucleation and heterogeneous nucleation processes can be thoroughly quantified by determining the nucleation rate, which is defined as the number of mature nuclei created per unit volume time in the system. The nucleation rate is usually highly correlated to many factors, especially the concentration of the nucleating phase and the surrounding chemical environments.

### 6.5.2.1 Homogeneous Nucleation

According to classical nucleation theory, nucleation occurs only if the nucleation barriers can be overcome [62, 63]. In addition, the newly formed crystalline phases are not thermodynamically stable until the dimensions are beyond the critical size. Mathematically, the nucleation barrier  $\Delta G^*$  and critical size  $R_c$  are represented by Eq. 6.6.

$$\Delta G^* = \frac{16\pi\gamma^3}{3(\rho_c\Delta\mu)^2} \text{ and } R_c = \frac{2\gamma}{\rho_c\Delta\mu} \quad (6.6)$$

Here,  $\rho_c$  is the particle density in the nuclei,  $\gamma$  is the surface free energy area density, and  $\Delta\mu$  is the chemical potential difference between the mother and crystalline phases [62]. As observed in Eq. 6.1, it is evident that  $\Delta\mu$  is the prime factor; hence, we need to highlight and comprehensively discuss it. The following Eq. 6.7 can be obtained from the definition of  $\Delta\mu$ .

$$\frac{\Delta\mu}{k_B T} = \ln \frac{a}{a^{eq}} \approx \ln \frac{C}{C^{eq}} \quad (6.7)$$

Here,  $a^{eq}$  and  $C^{eq}$  represent the equilibrium activity and concentration, respectively. By altering either the concentration of SF solutions or the equilibrium concentration for SF molecules (can be achieved by changing the pH, ionic strength, temperature, and other properties), the crystallization kinetics can be effectively controlled.

As the probability of nucleation is uniform throughout the entire system, the above process is known as homogeneous nucleation. In general, it is difficult for homogeneous nucleation to occur because of its extremely high nucleation barrier.

### 6.5.2.2 Heterogeneous Nucleation

In real nucleation cases, foreign bodies and substrates are always present in the system (e.g., the wall of solution containers, foreign particles, and substrates). In principle, if strong interactions can occur between the crystalline phase and foreign substrates, the occurrence of such foreign bodies will significantly lower the nucleation barrier; hence, the probability of the occurrence of nucleation adjacent to the foreign bodies is higher than elsewhere in the system. This is referred to as heterogeneous nucleation [16, 62]. To quantify the ability of foreign bodies to lower the nucleation barrier with regard to the homogeneous nucleation barrier, the interfacial correlation factor  $f(m)$  has been proposed. The parameter  $m$  describes the structural match between the crystalline phase and the substrate. Specifically, in the case of a perfect match,  $f(m) \sim 0$ . This implies that the heterogeneous nucleation barrier vanishes completely when the nucleating phase is well ordered and oriented along the structure of the foreign body substrates. However, when the structural match is poor (i.e.,  $f(m) \rightarrow 1$ ), there is almost no correlation between the foreign body substrates and the nucleating phase. In this extreme case, the substrate has almost no influence on the nucleation process, which is equivalent to homogeneous nucleation [16, 62]. In nucleation cases,  $f$  usually ranges from 0 to 1, which suggests that primary nucleation is somehow governed by heterogeneous nucleation.

Specifically, in the case of heterogeneous nucleation in SF molecules, the role of foreign substrates/surfaces/nanoparticles in directing the self-assembly of SF proteins and polypeptides has recently been reported. By carefully selecting functional nanomaterials and controlling the conditions of the SF solution, the foreign substrates can show a strong templating effect during SF protein heterogeneous nucleation. For instance, some mono-dispersed polystyrene nanoparticles (PS NPs),

which are coated with numerous functional carboxyl groups on the surface, can serve as ideal foreign bodies that provide nucleation sites. At these nucleation sites, the interactions between such nanoparticles and SF molecules mainly include the formation of H-bonds between the amino groups of SF peptide chains and carboxyl groups on the nanoparticle surfaces [19]. This strong interaction can reduce the nucleation barrier and accelerate the nucleation of  $\beta$ -sheets as well as  $\beta$ -crystallites.

It is worth mentioning that the template effect of the foreign body is not rare. For instance, graphene is also reported to be able to trigger SF heterogeneous crystallization [64]. By precisely controlling experimental conditions, the almost complete coverage of graphene nanosheets can be achieved by layers of densely packed SF nanofibrils [64]. In contrast, very few ( $\sim 1\%$ ) SF nanofibrils are observed outside of graphene nanosheets, suggesting that nanofibril formation (heterogeneous nucleation) occurs in a highly selective manner, only on graphene nanosheets [64]. This heterogeneous nucleation of SF molecules on foreign bodies might shed new light on the synthesis of composite silk materials.

We notice that this nucleation-controlled SF network formation model allows for the interpretation of many novel effects observed in SF materials. Owing to the nature of nanocrystallites, the formation of nanofibrils (crystal networks) is controlled by inter-molecular nucleation, which allows for the reconstruction and meso-functionalization of SF materials.

### 6.5.3 Experiments on SF Nucleation

#### 6.5.3.1 Nucleation Kinetics

As discussed, the pathway through which SF molecules self-assemble into higher levels of structures is controlled by a nucleation mechanism. According to the nucleation theories of soft material formation, nucleation can be divided into two types: homogeneous and heterogeneous, which determine nucleation rate  $J$  by Eq. 6.8:

$$J = A \exp[-\Delta G^* f / kT] \times N^0 \quad (6.8)$$

where  $A$  and  $B$  are kinetic parameters,  $f$  is the interaction parameter between the nucleating phase and templates ( $0 < f \leq 1$ , in the case of homogeneous nucleation,  $f = 1$ ).  $\Delta G^*$  is the nucleation barrier, and  $\Delta\mu$  is the chemical potential difference between the mother and crystalline phases.

In the process of nucleation, the addition of appropriate nucleation templates/seeds to the SF solution lowers the nucleation barrier to promote the nucleation rate. Equation (6.8) indicates that under the same experimental conditions, the nucleation rate  $J$  is directly proportional to the density of the added nucleation seeds  $N^0$ .

In principle, it is difficult to measure the nucleation rate directly. For this reason, nucleation rates are always compared by measuring nucleation induction time  $\tau$ . The basis of this alternative method lies in the fact that the occurrence of  $\beta$ -crystallites will

be followed by the formation of SF nanofibrils/ $\beta$ -crystallite networks, which consequently drives the increase in storage modulus  $G'$  and the turbidity of SF solutions. In other words, the occurrence of nanofibrils/ $\beta$ -crystallite networks will lead to the onset of  $G'$  at  $t_g$ , which corresponds to the dynamic induction time of the nucleation of  $\beta$ -crystallite or nanofibril/ $\beta$ -crystallite networks, or to the onset of turbidity at  $t'_g$ , which corresponds to the static induction time of the nucleation of  $\beta$ -crystallite and nanofibril/ $\beta$ -crystallite networks. We have then  $eJ \sim 1/t_g \sim 1/t'_g$ .

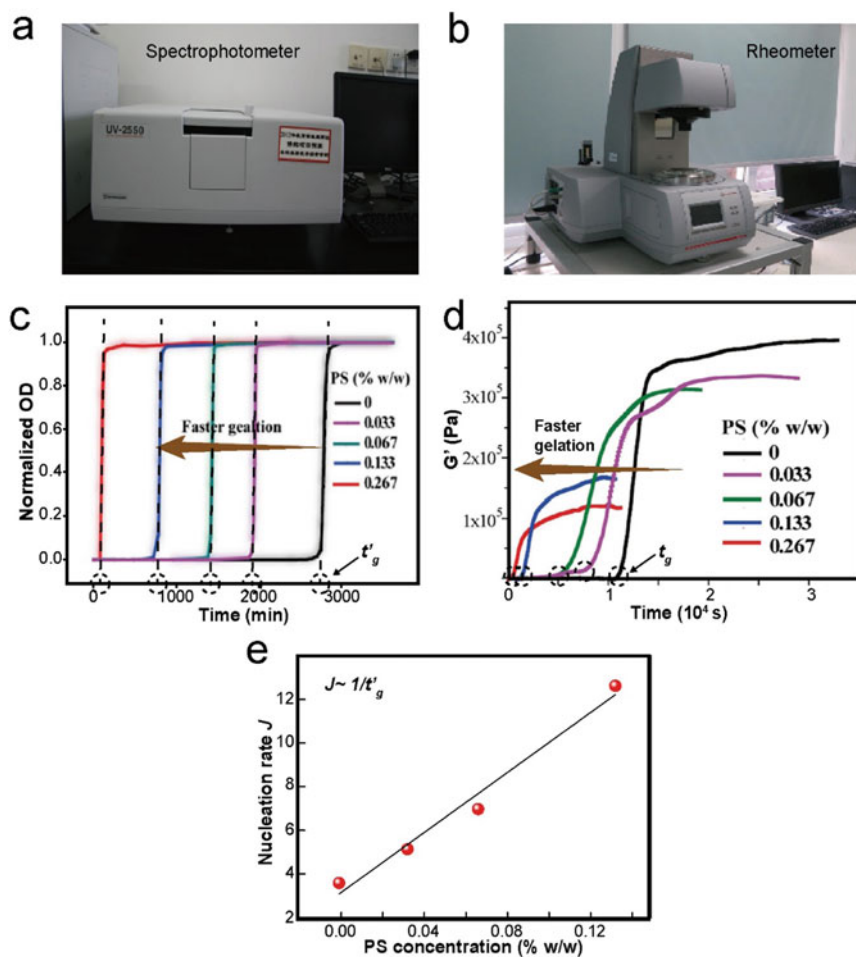
### 6.5.3.2 Experimental Verification of Nucleation Kinetics and Impact of Seeds

Taking  $\beta$ -sheets as a typical example of secondary structure, during CD measurements, the  $\beta$ -sheet nucleation rate can be measured by recording the temporal evolution of the 217 nm peak value (which is indicative of  $\beta$ -sheets) of either pure or composite RSF solutions. Afterwards, the value–time curve can be plotted together for further qualitative comparison. Given the fact that conformational transition from the random coil/ $\alpha$ -helix takes time, in a typical curve, the content of the  $\beta$ -sheet will initially remain nearly constant and then increase sharply. This deflection point is denoted as nucleation induction time  $\tau$  of the  $\beta$ -sheet.

To measure the increments of storage modulus  $G'$  and the turbidity of SF solutions, the characterization techniques of rheometry and spectrophotometry were applied, respectively. Figure 6.24 displays two typical examples that record the temporal evolution of the storage modulus and optical density (which correlates to the turbidity) of composite RSF solutions that are incorporated with different ratios of foreign templates (i.e., PS NPs). The shorter nucleation induction time is highly consistent with the accelerated nucleation kinetics.

Based on Eq. (6.3), under the same experimental conditions, one should have a linear relationship between  $J$  ( $\sim 1/t_g$ ) and the density of nucleation templators/seed, if the templators/seeds can promote nucleation. Given the relatively large dimensions of SF nanofibrils (20–50 nm) and nanofibril networks, it is possible to investigate SF nucleation kinetics by directly monitoring the SF molecule accumulation process via advanced microscopy techniques such as SEM. In addition, by labeling SF molecules with specific fluorescent dyes, confocal laser scanning microscopy can also be used to record the formation rates of nanofibrils. For instance, Chen et al. studied the promotion effect of foreign nanoparticles (PS NPs) on SF nanofibril formation, and their results clearly showed that green fluorescence-labeled SF molecules accumulated around the surface of PS NPs. In addition, the SEM images of the nanoparticles incubated in SF solutions also clearly displayed the accumulation of SF nanofibrils on their surface; with a longer induction time, the number of SF nanofibrils increased.

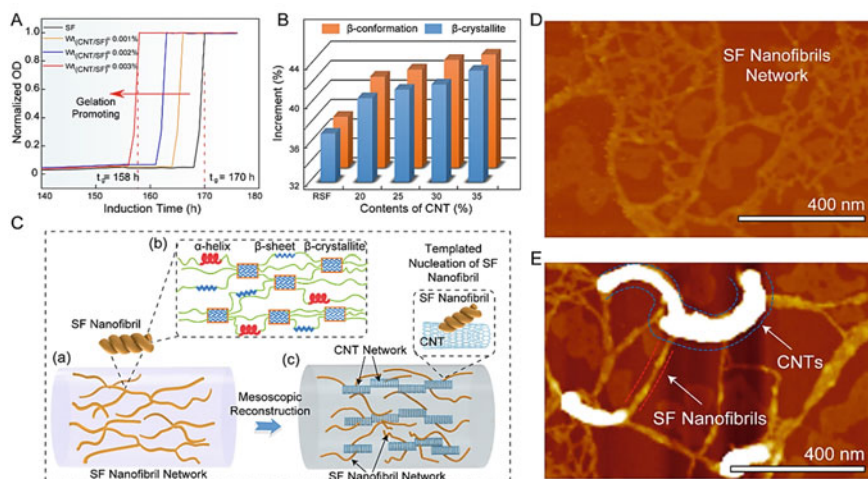
In a similar manner to PS NPs, it has been reported that CNTs are also capable of accelerating the gelation process of SF solution, which has been confirmed by their reduced gelation time (or nucleation induction period). Figure 6.25a clearly shows that the gelation time decreased with increasing CNT content. Nevertheless, the influence of CNTs on the content of secondary and tertiary structures within SF



**Fig. 6.24** Typical techniques measuring nucleation rates **a** Image of spectrophotometer applied to measure turbidity of solution. **b** Image of rheometer applied to record temporal evolution of storage modulus of solution. **c** Normalized OD changes of neat SF solution and SF-PS mixtures with various PS concentrations. **d** Storage moduli of SF-PS mixtures at various PS concentrations as a function of time. Reproduced with permission [19]. Copyright 2019, Wiley-VCH. **e** Linear correlation between nucleation rate  $J$  and the density of the added nucleation seeds  $N^\circ$

materials was investigated by FTIR and WAXD, respectively. The results show that both  $\beta$ -conformations (including both  $\beta$ -sheets and  $\beta$ -crystallites) and  $\beta$ -crystallites increased constantly with increasing CNT content in the CNT/SF composite fibers, further indicating that CNTs are favorable for the nucleation process of  $\beta$ -sheets and  $\beta$ -crystallites.

Figure 6.25c displays the mesoscopic hierarchical networks of neat SF fibers, in which the most important structural units are interpenetrating SF nanofibrils. When



**Fig. 6.25** Interpretation of gelation process of silk (corresponding formation of SF nanofibrils networks), effect of additive carbon nanotubes (CNTs), and images of nanofibrils of pure SF and CNT/SF composited fibers. **A** Gelation kinetics of pure SF and mixed CNT/SF solution. **B** Influence of CNT content on the content of  $\beta$ -conformation and  $\beta$ -crystallites in CNT/SF fibers. **C** Flow chart illustrates reconstruction of mesoscopic network: (a) pure SF nanofibril networks, (b) illustration of secondary structure within nanofibril, and (c) hybrid CNT/SF mesoscopic networks. **D–E** AFM images of **D** pure and **E** composite SF nanofibril networks. Reproduced with permission [17]. Copyright 2020, Wiley-VCH

CNTs are added to a SF solution, the surface of the CNTs can serve as a foreign substrate for the promotion of the heterogeneous nucleation of SF nanofibril formation. The AFM images of pure SF nanofibril networks and CNT/SF networks are shown in Fig. 6.25d and e, respectively. It follows that due to the templating effect, CNTs were incorporated into the SF fibril networks. Owing to the reconstruction of SF fiber mesoscopic structures with additive CNTs (Fig. 6.25b), both the  $\beta$ -crystallites within the crystal network and the nodes (referring to the joints of CNT/SF fibrils) within the nanofibril network increased, which consequently strengthened the mechanical performance of the CNT/SF composite fibers.

## 6.5.4 Reconstruction and Meso-Functionalization of Silk Fibers

### 6.5.4.1 Natural Silk Fibers

*Genetic Modification to Produce Functionalized Silk Fibers:* As discussed, the primary structure of silk fibers refers to the amino acid sequences. If the amino acid sequences are changed, the corresponding structure and mechanical performance are

also altered. It is reported that genetic modification techniques have been successfully applied for producing fluorescent colored silks [65]. According to previous studies, the gene for the fluorescent colored protein (e.g. green fluorescent protein, GFP) can be inserted into the silkworm genome and then the SF proteins and fluorescent proteins are co-expressed simultaneously. Consequently, transgenic, fluorescent colored silk fibers are obtained. As the primary structure of silk fibers is incorporated within the sequence of fluorescent proteins, the mechanical strength of the hybrid silk fibers is slightly improved (or decreased) [65].

Although fluorescent protein sequences can be incorporated into the primary structure of silk fibers and consequently change the mechanical properties, the modulation ability is not satisfying. Because the mechanical properties of spider silks (especially the dragline silk fiber) are superior to those of silkworm silk, numerous efforts have recently been devoted to the use of genetically modified silkworms as the host for the production of transgenic spider silk fibers [66, 67]. Previously, most relevant attempts focus on transposon-mediated transgenic silkworms. These silkworms can successfully produce reinforced silk fibers. However, the amount of spider silk proteins in these transgenic silkworms is very low (<5%); this may be attributable to the variable promoter activities and endogenous SF protein expression [67]. Recently, Xu et al. [66] put forward a system for the production of massive amounts of spider silk in silkworms using a transcription activator-like effector nuclease (TALENs)-mediated technique, to replace the fibroin H-chain gene of silkworms with the *MA* spideroin-1 gene. According to their experimental results, the achieved yield of chimeric spider protein within the obtained hybrid silk fibers is up to 35.2% of the total amount. More importantly, the relative abundance of the primary structure of spideroin significantly improves the toughness of the transgenic fibers, and especially increase the extensibility [66].

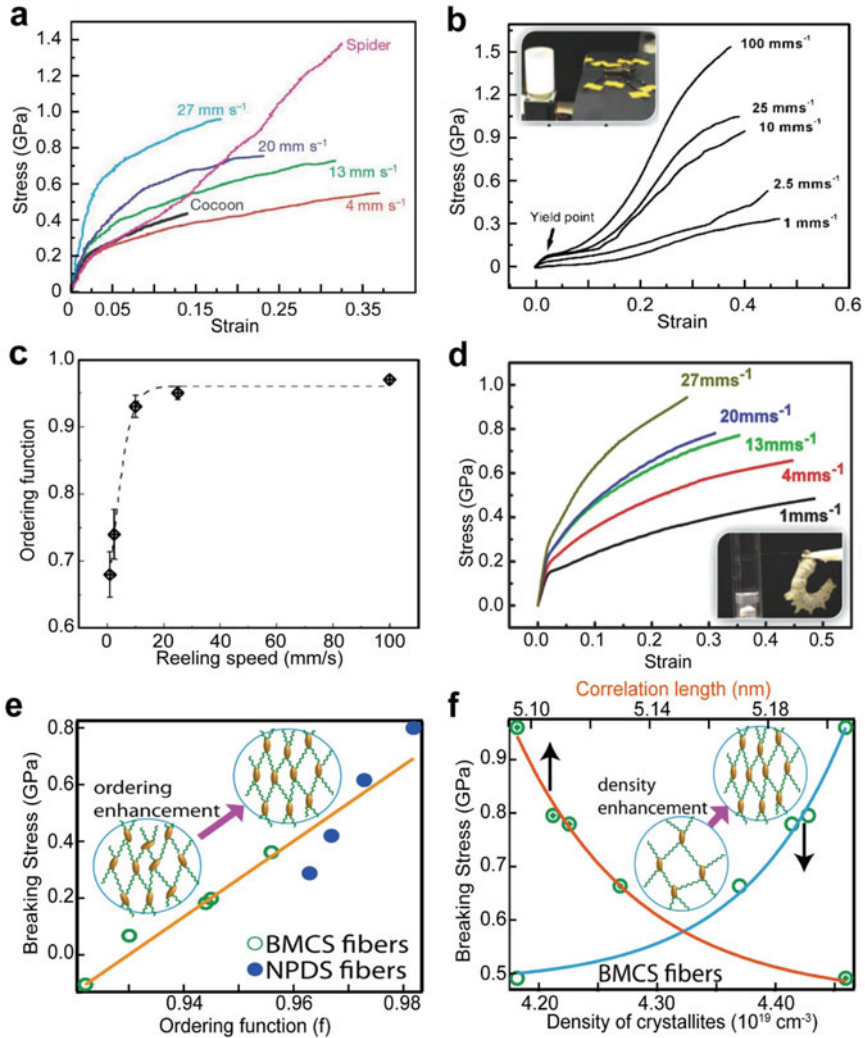
The feeding technique is another *in vivo* method for producing intrinsically functionalized silk fibers. In comparison to the gene modification method, the feeding method is much more versatile and effective. Besides, this method can also be well understood within the framework of SF molecule crystallization theory. It is reported that fluorescent silk cocoons and fibers can be directly obtained by simply feeding silkworms various small fluorescent molecules (e.g., rhodamine B, rhodamine 101, and rhodamine 110) [9]. It follows that all these small fluorescent molecules can directly conjugate with the SF molecules via molecular recognition [68]. Thus, these small fluorescent dyes are trapped into the silk fibers along with the SF molecule crystals during the natural spinning process. Further structural characterization confirms that the secondary structure of silk fibers is modified via feeding [68]. Specifically, micro-FTIR shows that there is no apparent structural homogeneity within silk fibers; such structural fluctuations might be the reason why these modified fibers display unsatisfactory mechanical properties (e.g., tensile strength) [68]. Apart from the small fluorescent dye molecules, it was recently reported that some specific nanoscale materials, such as functionalized TiO<sub>2</sub> nanoparticles [69], single-walled carbon nanotubes (SWNTs), and graphene [70], can also be directly conjugated into silk fibers by feeding silkworms artificial diets. Moreover, these additives can greatly improve the macroscopic performance of the composite silk fibers. For instance, it was determined



that a 1% ratio of TiO<sub>2</sub> nanoparticles greatly enhanced both the breaking strength and elongation of the composite fibers [69]. In addition, these functionalized TiO<sub>2</sub> particles can also endow the composite fibers with anti-ultraviolet properties; this is extremely important for the long-term storage of silk fibers [69]. Similarly, Wang et al. [70] obtained mechanically enhanced silk fibers by feeding *B. mori* larval silkworms SWNTs and graphene. In addition, the incorporation of graphene and SWNTs can increase the conductivity of these fibers [70].

To quantify the influence of additives, a structural survey was also carried out, particularly for the second level. Because the  $\alpha$ -helix and random coil are relatively softer than the  $\beta$ -sheet, a higher  $\alpha$ -helix/random coil structure content should result in a material with an improved breaking tensile strength and modules with greater toughness. However, the  $\beta$ -sheet content in the composite fibers was found to be relatively smaller than that of the natural control fibers, which suggests that the presence of SWNTs and graphene in the silk matrix may hinder the conformational transition from random coil/ $\alpha$ -helix to  $\beta$ -sheet [70]. To some degree, this result is reasonable because the relatively large dimensions of SWNTs and graphene might induce a steric hindrance effect and consequently prevent the crystallization of SF molecules. However, this finding conflicts with the hierarchical network structure model proposed in this chapter, which claims that a lower  $\beta$ -sheet content should be accompanied by smaller breaking stress. Nevertheless, the enhancement of breaking strength is reportedly considerably attributable to the presence of relatively stronger mechanical additives (i.e., the SWNTs and graphene) [70]. Specifically, it is hypothesized that these additives in the silk matrix may act as knots (similar slipknot structures are widely found in other biological structures such as proteins and DNA strands) [71]. They may also act as a key frictional element, reshaping the entire fiber and dissipating the additional fracture energy, which consequently results in enhanced mechanical properties in the fibers. The inclusion of additional additives in the composite fibers may not necessarily produce stronger fibers; on the contrary, it is determined that if excessive additives are incorporated into the silk matrix, such additives tend to aggregate and cause defects, which eventually results in silk fibers with inferior mechanical performance [69, 70]. In summary, this natural feeding method is not only straightforward but also effective. More importantly, this method can be scaled up easily and can thus shed light on the massive production of reinforced silk fibers.

*Forced Reeling of Natural Silk Fiber:* As discussed, with an increase in the force reeling speed of both silkworm silk (Fig. 6.26a and d) and spider MA fibers (Fig. 6.26b), the mechanical performance, and especially the breaking stress values, can be improved accordingly. Structural characterization reveals that it is the meso-reconstruction of the hierarchical structures, especially the level 3 and level 4 structures, that results in mechanical enhancement. Specifically, the crystallite size decrease monotonously with the acceleration of the reeling speed. Meanwhile, the orientation function and the density of crystallites increased (Fig. 6.26c) [18]. Recently, Xu et al. [32] have revealed that the key structural parameters, i.e.,  $f$  (orientation function),  $n_{\beta}$  (the number of crystallites at the cross section of a nanofibril),



**Fig. 6.26** Influence of Forced Reeling Speed on Mechanical Performance of Fibers and Orientation Function. **a** Silkworm silk fibers spun by forced reeling can be compared to spider fibers. Reproduced with permission [6]. Copyright 2002, Springer Nature; **b** spider *MA* dragline fibers under different reeling speeds. Reproduced with permission [34]. Copyright 2015, Wiley-VCH. **c** Orientation function of the force-spun silk fibers in **(b)** increases with the reeling speed. Reproduced with permission [8]. Copyright 2006, Elsevier. **d** Silkworm silk fibers under different reeling speeds. Reproduced with permission [34]. Copyright 2015, Wiley-VCH. **e** the influence of the ordering function of crystallites; the breaking stress of both BMCS fibers and NPDS fibers increases linearly with an increase in the value of  $f$ . Both fibers can be fitted by the same curve, which suggests that the strength of the two types of fibers is governed by similar structures, i.e. the hierarchical network structure. **f** The influence of the density and correlation length of the crystallites was observed; the breaking stress of both BMCS fibers increases exponentially with an increase in the density of crystallites but decreases exponentially with the correlation length. **e, f** Reproduced with permission [18]. Copyright 2016, Wiley-VCH

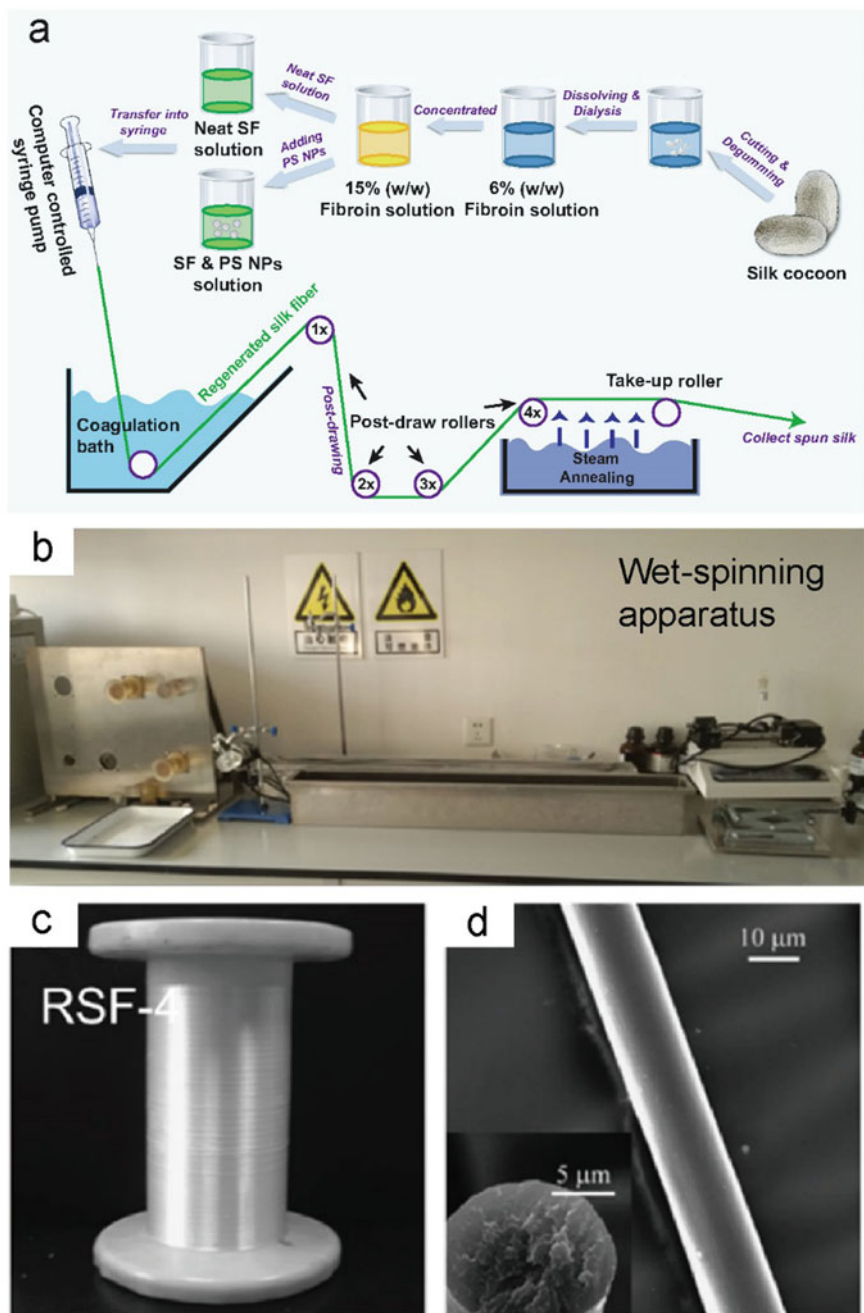
A (effective loading area of a peptide chain in the  $\beta$ -crystallites), etc. were changed because of the reeling speed, consequently resulting in modifications in the toughness of silk fibers. In a similar way, Liu et al. [18] also measured the orientation function and density of crystallites with varying the reeling speed, and quantitatively studied the correlation with the mechanical performance of silk fibers (Fig. 6.26e and f).

#### 6.5.4.2 Artificial Spinning of Regenerated Silk Fibroin (RSF) Fiber Based on Templated Nucleation

Although the above techniques can successfully produce mechanically enhanced natural silk fibers (neat or composite), the mechanical properties of the obtained silk fibers still greatly depend on the rearing conditions of the silkworms. With the development of techno-polymer fibers based on petrochemicals, increasing efforts have been devoted to artificially extruding strong and tough silk fibers from regenerated silk protein solutions. So far, electro-spinning, dry-spinning, and wet-spinning techniques have been developed to successfully fabricate mechanically enhanced RSF fibers. Among these techniques, dry-spinning and wet-spinning techniques greatly bio-mimic the natural silk fiber reeling process *in vitro*, in which natural fibers are produced under benign and physiological conditions (i.e., at ambient temperature and relatively low hydraulic pressures using water as the protein solvent). In particular, the wet-spinning technique, which is performed via the ejection of the spinning dope solution into a specific coagulation bath, is the most common and versatile approach; hence, we will comprehensively introduce it within the framework of the SF molecule crystallization theory in this subsection.

In general, there are two different methods for the reinforcement of artificial RSF fibers. One involves the control of the spinning process to promote homogeneous SF nucleation kinetics in which neat SF solutions are applied as the feedstock. The other method involves the addition of specific reinforcement agents or foreign nanoparticles into the dope solution to extrude composite artificial RSF fibers. By activating the heterogeneous nucleation of the SF molecules, the nucleation barrier can be suppressed, thereby increasing the probability of crystallization, which consequently results in a higher crystal density and smaller correlation length of the molecular crystal networks within the composite RSF fibers. Thus, mechanically enhanced RSF fibers can be synthesized. To date, this method has been successfully used to fabricate several significantly reinforced composite RSF fibers (or mats).

Figure 6.27 displays a typical procedure and the corresponding experimental instruments for wet spinning artificial RSF fibers. Prior to wet spinning, the SF solution was mixed with certain ratios of specific foreign templates and was then stored for 2 h. Next, the composite RSF solutions were transferred to a steel syringe, and a high-pressure injection pump was utilized to extrude the solution into a coagulation solution (e.g., aqueous 35% (w/v)  $(\text{NH}_4)_2\text{SO}_4$  solution) at 20 °C. The fibers were wound from the coagulation bath over four rollers at an increasing rotating speed. As a result, a larger external shear force was applied to the as-obtained RSF



**Fig. 6.27** **a** Schematic of experimental procedures of neat SF and hybrid fibers and **b** images of wet-spinning apparatus. **c** Optical images of the final-obtained artificial RSF fiber. **d** SEM images of morphology and crosssection of the artificial RSF fiber

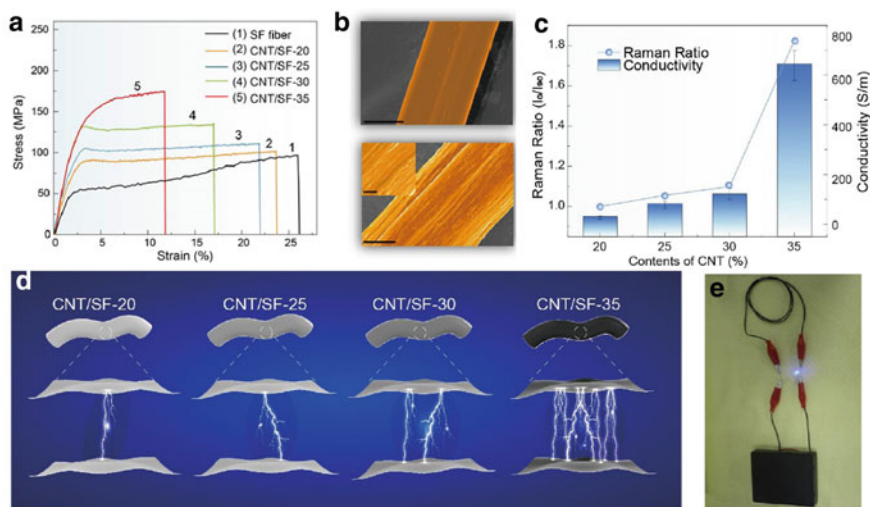
fibers to promote SF crystallization and, consequently, to improve the crystallization as well as the mechanical performance. Finally, an additional steam-annealing treatment was applied to enhance the crystallinity and mechanical properties of the fibers. The final artificial RSF fibers were obtained by the take-up device.

Chen et al. [19] introduced PS NPs into an RSF solution and then utilized a homemade apparatus to synthesize RSF composite fibers. By simply varying the PS NPs content added, the macroscopic performance of the final obtained composite RSF fibers was altered due to their strong interactions with SF molecules. Consequently, the addition of PS NPs can greatly improve mechanical performance [19]. Within the framework of the continuous nucleation mechanism (cf. Fig. 6.23d, e), this affinity interaction between PS NPs and SF molecules (specifically referred to as the intermolecular hydrogen bonds between the carboxyl groups on the surface of PS NPs and amino groups within SF peptide chains) can significantly decrease surface tension and lower the nucleation barrier. Thus, the heterogeneous nucleation of SF molecules is greatly promoted on/near the surface of PS NPs, which results in the generation of more nucleation sites (corresponding to a larger crystal density within the spun RSF fibers), and alteration of crystal networks, and the structures of nanofibrils. This gives rise to the reconstruction of mesoscopic hierarchical structures of SF materials (cf. Fig. 6.23). As discussed, the crystallization in both natural silk fibers and RSF fibers exhibit strong structural interactions. Consequently, the increase in density will result in stronger fibers.

Apart from the aforementioned nanoparticles, other functionalized composite RSF fibers (or mats), such as those containing hydroxyapatite (HAP) nanoparticles [72] and cellulose [73] have also been successfully fabricated using the electrospinning technique.

However, it is notable that not all additives are beneficial for the heterogeneous nucleation of SF molecules [74]. In fact, some additives have a high tendency for self-assembly. When such additives are mixed with SF dope solutions, large clusters (or aggregates) of additives will form spontaneously, which leads to defects and results in the generation of fibers that are inferior to composite RSF fibers. Hence, a uniformly dispersed mixture solution is essential. Normally, small ratios of surfactants can be incorporated into the mixture to increase the dispersion of additives. Upon performing this procedure, the obtained composite RSF fibers generally display a bead-free nanofiber morphology that does not exhibit any aggregation of additives [72] and the mechanical properties are improved. Further structural analyses have clearly shown that such composite RSF fibers have a larger  $\beta$ -sheet and  $\beta$ -crystallite content than the neat RSF fibers [74].

Another notable finding is that the addition other functionalized additives into SF materials during the formation is an effective method for endowing RSF fibers with some new and unique performance. Such added performance can range from the biological, optical to electronic aspects [72–74]. For instance, Ma et al. successfully functionalized SF fibers by carbon nanotubes based on the CNT-templated nucleation of SF networks. This leads to electrically conductive biocompatible RSF fibers. In this context, both the mechanical properties and the electronic conductivity of the reconstructed mesoscopic functional fibers can be tuned by varying the density of the



**Fig. 6.28** Comparison of strength, morphologies and conductive performance of neat and CNT/SF hybrid fibers, and an illustration of conductive principle. **a** Stress–strain curves of SF and CNT/SF containing specific quantities of CNTs (0, 20, 25, 30, and 35 wt%). **b** SEM image of typical surface morphology of SF fibers (Scale bar in the image is 10  $\mu\text{m}$ .) and SEM image of typical surface morphology of CNT/SF-35 fiber (Scale bar inside the image is 1  $\mu\text{m}$ , and the bar outside is 10  $\mu\text{m}$ .) **c** Polarized Raman ratio, and electrical conductivity of different CNT contents in CNT/SF fibers (20%, 25%, 30% and 35% (wt%)). **d** Schematic illustration of percolation of CNT/SF hybrid fibers. Conductive channels continuously increase with the integration of CNTs and experience a sharp increase in content of CNT/SF-35 >30 wt%, which can be attributed to percolation of CNTs in CNT/SF hybrid fibers, which is when aggregation of CNTs in SF hybrid occurs. **e** CNT/SF yarn can be used as conductive wire. Reproduced with permission [17]. Copyright 2020, Wiley-VCH

incorporated CNTs. In particular, when the content of the CNTs exceeded 35 wt%, the conductivity reached 638.9 S/m, which is eightfold higher than the best available materials of similar types (Fig. 6.28).

## 6.6 Conclusions and Perspectives

In this chapter, the hierarchical network structures and the correlations between the structure and performance of SF materials have been extensively discussed. In a bottom-top manner, the hierarchical network structure is identified as having five levels, that is, the amino acid sequence, secondary structure,  $\beta$ -crystallites, crystal network and nanofibril network. It follows that the outstanding mechanical performance of SF materials synergistically results from the nano-fishnet topology structure of  $\beta$ -crystallites in the molecular-scale crystal network and from the strong linkage (friction) among nanofibrils in the mesoscopic nanofibril network. In comparison, although non-fibrous SF materials (i.e., films, hydrogels, and scaffolds) and silk

fibers share similar hierarchical structures, the difference in inter-nanofibril interaction strength and nanofibril architecture results in distinct performance. Specifically, within the framework of the crystallization of SF molecules and network formation mechanism of SF materials, an increase in the crystallinity and a better alignment of  $\beta$ -crystallites will correspondingly result in stronger materials. The knowledge obtained in this chapter will shed light on the preparation of ultra-high-performance SF materials from crystallization and structural points of view.

## References

1. Vepari, C., Kaplan, D.L.: *Prog. Polym. Sci.* **32**(8–9), 991 (2007)
2. R. F. Service: *Science* **322**, 1460 (2008)
3. Zhu, B., Wang, H., Leow, W.R., Cai, Y., Loh, X.J., Han, M.Y., Chen, X.: *Adv. Mater.* **28**(22), 4250 (2016)
4. Hu, F., Lin, N., Liu, X.Y.: *iScience* **23** (4), 101035 (2020)
5. Shi, C., Wang, J., Sushko, M.L., Qiu, W., Yan, X., Liu, X.Y.: *Adv. Funct. Mater.* **29** (42) (2019)
6. Shao, Z.Z., Vollrath, F.: *Nature* **418**(6899), 741 (2002)
7. Vollrath, F., Knight, D.P.: *Nature* **410** (6828), 541 (2001); Omenetto, F.G., Kaplan, D.L.: *Science* **329** (5991), 528 (2010)
8. Du, N., Liu, X.Y., Narayanan, J., Li, L.A., Lim, M.L.M., Li, D.Q.: *Biophys. J.* **91**(12), 4528 (2006)
9. Tansil, N.C., Li, Y., Teng, C.P., Zhang, S.Y., Win, K.Y., Chen, X., Liu, X.Y., Han, M.Y.: *Adv. Mater.* **23**(12), 1463 (2011)
10. Girotti, A., Orbanic, D., Ibanez-Fonseca, A., Gonzalez-Obeso, C., Rodriguez-Cabello, J.C.: *Adv. Healthc. Mater.* **4**(16), 2423 (2015)
11. Tao, H., Kaplan, D.L., Omenetto, F.G.: *Adv. Mater.* **24**(21), 2824 (2012)
12. Abelein, A., Abrahams, J.P., Danielsson, J., Graslund, A., Jarvet, J., Luo, J.H., Tiiman, A., Warmlander, S.K.T.S.: *J. Biol. Inorg. Chem.* **19**(4–5), 623 (2014)
13. Tian, Y., Jiang, X.J., Chen, X., Shao, Z.Z., Yang, W.L.: *Adv. Mater.* **26**(43), 7393 (2014)
14. Zourob, M., Gough, J.E., Ulijn, R.V.: *Adv. Mater.* **18**(5), 655 (2006)
15. Wang, M.T., Braun, H.G., Kratzmuller, T., Meyer, E.: *Adv. Mater.* **13**(17), 1312 (2001)
16. Lin, N.B., Liu, X.Y.: *Chem. Soc. Rev.* **44**(21), 7917 (2015)
17. Ma, L., Liu, Q., Wu, R., Meng, Z., Patil, A., Yu, R., Yang, Y., Zhu, S., Fan, X., Hou, C., Li, Y., Qiu, W., Huang, L., Wang, J., Lin, N., Wan, Y., Hu, J., Liu, X.Y.: *Small* **16**(26), e2000203 (2020)
18. Liu, R.C., Deng, Q.Q., Yang, Z., Yang, D.W., Han, M.Y., Liu, X.Y.: *Adv. Funct. Mater.* **26**(30), 5534 (2016)
19. Chen, Z.W., Zhang, H.H., Lin, Z.F., Lin, Y.H., van Esch, J.H., Liu, X.Y.: *Adv. Funct. Mater.* **26**(48), 8978 (2016)
20. Zelzer, M., Ulijn, R.V.: *Chem. Soc. Rev.* **39**(9), 3351 (2010)
21. Kundu, S.C., Kundu, B., Talukdar, S., Bano, S., Nayak, S., Kundu, J., Mandal, B.B., Bhardwaj, N., Botlagunta, M., Dash, B.C., Acharya, C., Ghosh, A.K.: *Biopolymers* **97**(6), 455 (2012)
22. Koh, L.D., Cheng, Y., Teng, C.P., Khin, Y.W., Loh, X.J., Tee, S.Y., Low, M., Ye, E.Y., Yu, H.D., Zhang, Y.W., Han, M.Y.: *Prog. Polym. Sci.* **46**, 86 (2015)
23. Li, X., Liu, Y., Zhang, J., You, R., Qu, J., Li, M.: *Mater. Sci. Eng. C Mater. Biol. Appl.* **72**, 394 (2017); Li, L., Puhl, S., Meinel, L., Germershaus, O.: *Biomaterials* **35** (27), 7929 (2014); Rockwood, D.N., Gil, E.S., Park, S.H., Kluge, J.A., Grayson, W., Bhumiratana, S., Rajkhowa, R., Wang, X., Kim, S.J., Vunjak-Novakovic, G., Kaplan, D.L.: *Acta Biomater.* **7** (1), 144 (2011); Luo, Z., Li, J., Qu, J., Sheng, W., Yang, J., Li, M.: *J. Mater. Chem. B* **7** (1), 80 (2019)

24. Mitropoulos, A.N., Perotto, G., Kim, S., Marelli, B., Kaplan, D.L., Omenetto, F.G.: *Adv. Mater.* **26**, 1105 (2014)
25. Luetchford, K., Chaudhuri, J., De Bank, P.: *Eur. Cells Mater.* **23**, 70 (2012)
26. Jin, H.J., Kaplan, D.L.: *Nature* **424** (6952), 1057 (2003); Hagn, F., Eisoldt, L., Hardy, J.G., Vendrely, C., Coles, M., Scheibel, T., Kessler, H.: *Nature* **465** (7295), 239 (2010)
27. Brown, C.P., Harnagea, C., Gill, H.S., Price, A.J., Traversa, E., Licoccia, S., Rosei, F.: *ACS Nano* **6**(3), 1961 (2012)
28. Simmons, A.H., Michal, C.A., Jelinski, L.W.: *Science* **271**(5245), 84 (1996)
29. Vollrath, F., Holtet, T., Thogersen, H.C., Frische, S.: *Proc. Roy. Soc. B-Biol. Sci.* **263** (1367), 147 (1996); Putthanarat, S., Stribeck, N., Fossey, S.A., Eby, R.K., Adams, W.W.: *Polymer* **41** (21), 7735 (2000)
30. Termonia, Y.: *Macromolecules* **27**(25), 7378 (1994)
31. Knowles, T.P., Fitzpatrick, A.W., Meehan, S., Mott, H.R., Vendruscolo, M., Dobson, C.M., Welland, M.E.: *Science* **318** (5858), 1900 (2007); Lin, Z., Huang, W.D., Zhang, J.F., Fan, J.S., Yang, D.W.: *Proc. Natl. Acad. Sci. U. S. A.* **106** (22), 8906 (2009)
32. Xu, G.Q., Gong, L., Yang, Z., Liu, X.Y.: *Soft Matter* **10**(13), 2116 (2014)
33. Du, N., Yang, Z., Liu, X.Y., Li, Y., Xu, H.Y.: *Adv. Funct. Mater.* **21**(4), 772 (2011)
34. Nguyen, A.T., Huang, Q.L., Yang, Z., Lin, N.B., Xu, G.Q., Liu, X.Y.: *Small* **11**(9–10), 1039 (2015)
35. Karsai, A., Martonfalvi, Z., Nagy, A., Grama, L., Penke, B., Kellermayer, M.S.Z.: *J. Struct. Biol.* **155**(2), 316 (2006)
36. Oroudjev, E., Soares, J., Arcidiacono, S., Thompson, J.B., Fossey, S.A., Hansma, H.G.: *Proc. Natl. Acad. Sci. U. S. A.* **99**(14), 9606 (2002)
37. Qiu, W., Patil, A., Hu, F., Liu, X.Y.: *Small* **15**(51), e1903948 (2019)
38. Tu, H., Yu, R., Lin, Z.F., Zhang, L., Lin, N.B., Yu, W.D., Liu, X.Y.: *Adv. Funct. Mater.* **26**(48), 9032 (2016)
39. Zhou, C.Z., Confalonieri, F., Medina, N., Zivanovic, Y., Esnault, C., Yang, T., Jacquet, M., Janin, J., Duguet, M., Perasso, R., Li, Z.G.: *Nucleic Acids Res.* **28**(12), 2413 (2000)
40. Ha, S.W., Gracz, H.S., Tonelli, A.E., Hudson, S.M.: *Biomacromol* **6**(5), 2563 (2005)
41. Zhou, C.Z., Confalonieri, F., Jacquet, M., Perasso, R., Li, Z.G., Janin, J.: *Proteins-Struct. Funct. Genet.* **44** (2), 119 (2001); Takahashi, Y., Gehoh, M., Yuzuriha, K.: *Int. J. Biol. Macromol.* **24** (2–3), 127 (1999); Fossey, S.A., Nemethy, G., Gibson, K.D., Scheraga, H.A.: *Biopolymers* **31** (13), 1529 (1991)
42. Mendes, A.C., Baran, E.T., Reis, R.L., Azevedo, H.S.: *Wires Nanomed Nanobi* **5**(6), 582 (2013)
43. Guo, C.C., Zhang, J., Wang, X.G., Nguyen, A.T., Liu, X.Y., Kaplan, D.L.: *Small* **13** (47) (2017)
44. Wu, X., Liu, X.Y., Du, N., Xu, G.Q., Li, B.W.: *Appl. Phys. Lett.* **95** (9) (2009)
45. Ketten, S., Xu, Z.P., Ihle, B., Buehler, M.J.: *Nat. Mater.* **9**(4), 359 (2010)
46. Gong, Z.G., Huang, L., Yang, Y.H., Chen, X., Shao, Z.Z.: *Chem. Commun.* **48**, 7506 (2009)
47. Chen, X., Shao, Z.Z., Marinkovic, N.S., Miller, L.M., Zhou, P., Chance, M.R.: *Biophys. Chem.* **89**(1), 25 (2001)
48. Chen, X., Shao, Z.Z., Knight, D.P., Vollrath, F.: *Proteins-Struct. Funct. Bioinf.* **68**(1), 223 (2007)
49. Lefevre, T., Rousseau, M.E., Pezolet, M.: *Biophys. J.* **92**(8), 2885 (2007)
50. Rousseau, M.E., Lefevre, T., Beaulieu, L., Asakura, T., Pezolet, M.: *Biomacromol* **5**(6), 2247 (2004)
51. Yang, Y.H., Shao, Z.Z., Chen, X., Zhou, P.: *Biomacromol* **5**(3), 773 (2004)
52. Carrion-Vazquez, M., Oberhauser, A.F., Fisher, T.E., Marszalek, P.E., Li, H.B., Fernandez, J.M.: *Prog. Biophys. Mol. Bio.* **74**(1–2), 63 (2000)
53. Sinsawat, A., Putthanarat, S., Magoshi, Y., Pachter, R., Eby, R.K.: *Polymer* **43** (4), 1323 (2002); Trancik, J.E., Czernuszka, J.T., Bell, F.I., Viney, C.: *Polymer* **47** (15), 5633 (2006)
54. Shao, J.Z., Zheng, J.H., Liu, J.Q., Carr, C.M.: *J. Appl. Polym. Sci.* **96**(6), 1999 (2005)
55. Rousseau, M.E., Beaulieu, L., Lefevre, T., Paradis, J., Asakura, T., Pezolet, M.: *Biomacromol* **7**(9), 2512 (2006)
56. Drummy, L.F., Farmer, B.L., Naik, R.R.: *Soft Matter* **3**(7), 877 (2007)



57. Lefevre, T., Rousseau, M.E., Pezolet, M.: *Appl. Spectrosc.* **60**(8), 841 (2006)
58. Dicko, C., Knight, D., Kenney, J.M., Vollrath, F.: *Biomacromol* **5**(3), 758 (2004)
59. Canetti, M., Seves, A., Secundo, F., Vecchio, G.: *Biopolymers* **28**(9), 1613 (1989)
60. Li, G.Y., Zhou, P., Shao, Z.Z., Xie, X., Chen, X., Wang, H.H., Chunyu, L.J., Yu, T.Y.: *Eur. J. Biochem.* **268**(24), 6600 (2001)
61. Xing, Y., Shi, C., Zhao, J., Qiu, W., Wang, J., Yan, X.B., Yu, W.D., Liu, X.Y.: *Small* **13**(40), 1702390 (2017)
62. Zhang, T.H., Liu, X.Y.: *Chem. Soc. Rev.* **43**(7), 2324 (2014)
63. Zhang, K.Q., Liu, X.Y.: *Nature* **429** (6993), 739 (2004); Zhang, T.H., Liu, X.Y.: *J. Am. Chem. Soc.* **129** (44), 13520 (2007); Diao, Y.Y., Liu, X.Y.: *Adv. Funct. Mater.* **22** (7), 1354 (2012); Liu, X.Y.: In: *Low Molecular Mass Gelator*, p. 1. Springer (2005); Liu, X.Y.: In: *Advances in Crystal Growth Research*, p. 42. Elsevier (2001)
64. Ling, S.J., Li, C.X., Adamcik, J., Wang, S.H., Shao, Z.Z., Chen, X., Mezzenga, R.: *ACS Macro Lett.* **3**(2), 146 (2014)
65. Iizuka, T., Sezutsu, H., Tatematsu, K., Kobayashi, I., Yonemura, N., Uchino, K., Nakajima, K., Kojima, K., Takabayashi, C., Machii, H., Yamada, K., Kurihara, H., Asakura, T., Nakazawa, Y., Miyawaki, A., Karasawa, S., Kobayashi, H., Yamaguchi, J., Kuwabara, N., Nakamura, T., Yoshii, K., Tamura, T.: *Adv. Funct. Mater.* **23**(42), 5232 (2013)
66. Xu, J., Dong, Q.L., Yu, Y., Niu, B.L., Ji, D.F., Li, M.W., Huang, Y.P., Chen, X., Tan, A.J.: *Proc. Natl. Acad. Sci. U. S. A.* **115**(35), 8757 (2018)
67. Kuwana, Y., Sezutsu, H., Nakajima, K., Tamada Y., Kojima, K.: *PLoS One* **9** (8) (2014); Teule, F., Miao, Y.G., Sohn, Y.G., Kim, Y.S., Hull, J.J., Fraser, M.J., Lewis, R.V., Jarvis, D.L.: *Proc. Natl. Acad. Sci. U. S. A.* **109** (3), 923 (2012)
68. Li, K., Zhao, J.L., Zhang, J.J., Ji, J.Y., Ma, Y., Liu, X.Y., Xu, H.Y.: *ACS Biomater. Sci. Eng.* **1**(7), 494 (2015)
69. Cai, L.Y., Shao, H.L., Hu, X.C., Zhang, Y.P., Sustain, A.C.S.: *Chem. Eng.* **3**(10), 2551 (2015)
70. Wang, Q., Wang, C.Y., Zhang, M.C., Jian, M.Q., Zhang, Y.Y.: *Nano Lett.* **16**(10), 6695 (2016)
71. Pugno, N.M.: *PLoS One* **9** (4) (2014)
72. Kim, H., Che, L., Ha, Y., Ryu, W.: *Mat. Sci. Eng. C-Mater.* **40**, 324 (2014)
73. Zhou, L., Wang, Q., Wen, J.C., Chen, X., Shao, Z.Z.: *Polymer* **54**(18), 5035 (2013)
74. Pan, H., Zhang, Y.P., Hang, Y.C., Shao, H.L., Hu, X.C., Xu, Y.M., Feng, C.: *Biomacromol* **13**(9), 2859 (2012)

# Chapter 7

## A Primer on Gels (with an Emphasis on Molecular Gels)



Richard G. Weiss 

**Abstract** This chapter describes structural, kinetic, thermodynamic and viscoelastic aspects of how dispersions of small molecules (essentially 0-dimensional objects) aggregate into thermally reversible or thixotropic networks that immobilize large volumes of a liquid. When possible, the different properties of these molecular gels are correlated. The data are considered at different time and distance scales during the lives of the gels. A short history of molecular gels and challenges to advancing the field are also presented. The properties of some molecular gels with simple gelator structures, including long  $n$ -alkanes and derivatives of them, are described as well.

**Keywords** Hansen parameters · Schröder-van Laar equation · Ostwald ripening · Storage and loss modulus · Viscoelasticity · Critical gelator concentration · Avrami equation

### 7.1 Introduction: General Classifications

Gels are a part of almost all aspects of our materials world [1]. The processes by which gels form from particle structures are only one type of self-assembly. Others lead to a myriad of different phases which will not be discussed here, although each type depends on the relative magnitudes of enthalpic and entropic factors that are time-dependent [2]. Examples of hydrogels (i.e., gels based on aqueous liquid components) include aggregates of collagen (the most abundant protein in our bodies; when denatured, it can take the form of aspic or, as sold commercially with different additives in the US and elsewhere, Jell-O ©), amyloids (that have been linked to Alzheimer's disease), and other protein-based gels with actin, clathrin, and tubulin as the non-aqueous component. Other common hydrogels are in human and other mammalian bodies and jelly fish, and in jellied sand worms. Many types of toothpaste, deodorants, cosmetics, soaps, pharmaceutical delivery agents, and foods [3] are gels, as are many non-edible materials such as silly-putty, paints, inks, dental materials

---

R. G. Weiss (✉)

240 Reiss Science Building, Georgetown University, Washington, DC 20057-1227, USA  
e-mail: [weissr@georgetown.edu](mailto:weissr@georgetown.edu)

and other adhesives, materials for producing new morphs of pharmaceuticals [4], and aluminum soaps that gelate hydrocarbons for ‘fracturing’ in oil wells [5] and for producing napalm [6].

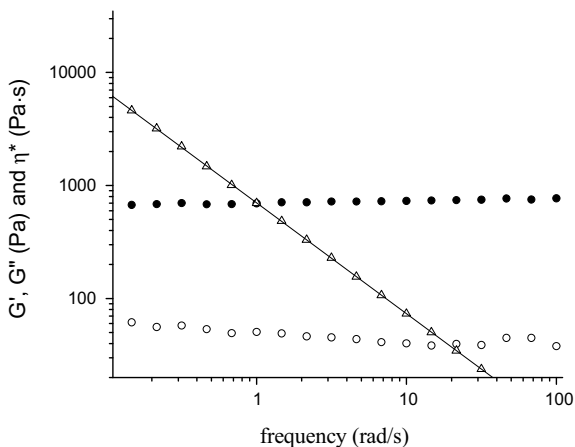
When the liquid components are aqueous or organic, the gels are referred to hydrogels or organogels, respectively. IUPAC defines a gel as “a non-fluid colloidal network or polymer network that is expanded throughout its whole volume by a fluid” [7]. This definition not very useful and is incomplete because it does not address the viscoelastic properties of gels. Also, it does not provide insights into the gel network structures of molecular gels and responses to stress at different length scales, from the macroscopic to the nanometric. Additionally, it avoids important questions concerning how adaptable are gel systems to changes in their component structures and environments. Only some of the factors listed below will be considered here:

- (1) How amenable is a class of molecular gelators to small structural modifications and what are the consequences of those changes to the properties of the gels?
- (2) How wide a range of liquid types can be gelated by a single gelator?
- (3) What is the lowest concentration of a molecular gelator that can gelate successfully a liquid at or near room temperature [i.e., the critical gelator concentration (CGC)]?
- (4) How long can a gel be maintained at room temperature without phase separation?
- (5) What is the temperature range over which a gel phase can be maintained? Specifically, at what temperature does a gel ‘melt’ ( $T_g$ ) and revert to its sol phase?
- (6) How strong or malleable is a particular gel? That is, what are its viscoelastic properties?

Common features of ‘gels’ are their compositions, at least two components—one of which is a solid and the other a liquid—their solid-like rheological behavior despite being mostly liquid [8], (i.e., they are non-Newtonian fluids), and their continuous microscopic structures with macroscopic dimensions that are permanent on the time scale of an analytical experiment [9]. Mechanical damping in gels is small:  $\tan\delta = G''/G' \ll 1$ , where  $G'$  is the storage modulus and  $G''$  is the loss modulus. The moduli are, respectively, measures of the energy stored and dissipated in a material in which a deformation has been imposed. Also, the oscillatory frequency dependence of the viscosity of a viscoelastic material under shear stress (i.e., the complex viscosity,  $\eta^* = \sqrt{(G'^2 + G''^2)/\omega}$ ) is a useful parameter to assess the viscoelasticity of a soft material. In fact, each of these is useful in assessing quantitatively whether a sample meets the viscoelastic criteria to be called a gel. A typical response of a gel to oscillatory frequency changes is shown in Fig. 7.1 [10]. Although cavitation rheology is a newer and potentially more useful method for assessing whether a sample is a gel [11], it has not been used extensively.

It is important to emphasize that microscopically the liquid component maintains many of its bulk viscosity and diffusion properties in the gel phase. The diffusion constant of the liquid or of a solute dissolved in the liquid portion is slower than in the absence of the gelator, but it is not inconsequential [12]. Only a small fraction

**Fig. 7.1** Typical frequency responses of the complex viscosity ( $\eta^*$ ,  $\Delta$ ), storage modulus ( $G'$ ,  $\bullet$ ) and loss modulus ( $G''$ ,  $\circ$ ) of a gel. Reprinted with permission from J Am Chem Soc 2006, 128, 15341. Copyright (2006) American Chemical Society



of the liquid molecules is in contact with the gelator network at any given time, and the liquid molecules are able to diffuse from one part of the gel to others. In addition, a constant fraction of the gelator molecules that are in the liquid phase (i.e., the aforementioned CGC) is also able to diffuse within the liquid component and exchange over time with molecules residing within the gelator networks. This equilibration and a thermodynamic preference for larger aggregates leads to Ostwald ripening over time that is observed in many gels. In fact, the liquid does not undergo macroscopic phase separation. Thus, destruction of a molecular gel, is principally a result of a lack of long-term balance between attractive and repulsive capillary and other interfacial forces; intrinsically, molecular gels are thermodynamically less stable than their phase-separated liquid and solid phases [13], although it may require years in some cases (and less than minutes, in others) to observe macroscopic phase separation. A simple semi-quantitative method for determining the CGC and  $T_g$  of gels is the ‘falling drop’ method, in which a metal ball is placed on the surface of gel and the concentration of gelator or temperature at which the ball falls is recorded [14].

As a result, identifying gels can be complicated because the name also encompasses a number of other ‘soft matter’ materials (e.g., microgels, colloids [15], emulsions, liquid crystals, and micelles) [2], each with different properties that generally do not meet all of the criteria noted as requisite in this chapter, although there are specific examples that do. Specifically excluded is ‘hard matter’ with inelastic networks, such as aluminosilicates. special formulations of some of the other soft materials may adopt all of the characteristics of and be gels. Of these, the general class of materials that are microgels are closest to ‘true’ gels: although lacking a continuous network that permeates the material, they do possess small gel-aggregates that are separated by a non-gel medium. An edible example of a microgel is bubble tea made from gelled seeds of *Hyptis suaveolens* (L.) Poir [16]. In addition, aerogels and xerogels are not gels although they are produced from them by removing the liquid component, leaving behind a solvent-free 3D or collapsed network.

## 7.2 A Short (Prejudiced) History of Gels

Although materials known to be gels have been mentioned for more than 3 millennia [17], the time span during which gels have been studied scientifically is much shorter. The author's prejudiced view of some of the most important observations and advances is summarized briefly below. In 1841, Lipowitz reported the first formal scientific molecular hydrogel 'sighting'—the gelation of aqueous solutions by lithium urate [18]. In 1864, Thomas Graham, who became Master of the Mint in England, began his studies of sol-gel chemistry; although these are not true gels as noted above, mention of this work is seminal because the pathway leading to them can involve 'true' gels. In addition, he made some very important (and unusual!!) pronouncements, such as his support for the theory of 'vitalism': "While the rigidity of the crystalline structure shuts out external expressions, the softness of the gelatinous colloid partakes of fluidity, and enables the colloid to become a medium for liquid diffusion, like water itself .... The colloid possesses *energia*. It may be looked upon as the probable primary source of the force appearing in the phenomena of vitality." [19].

In 1871, Maddox used 'dry' gelatin plates with silver salts for photography. In 1888, Eastman made silver halide dispersions in gelatin on cellulose nitrate rolls of film placed in a camera. In 1891, Meunier made gels with 1,3:2,4-di-O-benzylidene-D-sorbitol as the gelator. In 1896, Liesegang reported reactions of molecules, as well as their diffusion and crystallization, in gels. In 1899–1900, Hardy employed gel electrophoresis and recorded microscopic images of gel networks. In 1907, Foster and Jackson made organogels using camphoryl thioisemicarbazide as the gelator. In 1907, Cotton and Mouton performed studies of thixotropy in gels. In 1912, Zsigmondy and Bachmann reported gelation of aqueous and alcoholic liquids by fatty acid salts. Notably, in the same year, Hardy synthesized thermally-reversible gels with small organic molecules as the gelators and made serious attempts to formulate a theoretical basis for how rod-shaped objects aggregate into colloids and, in some cases, result in gels. Many discoveries have been made in the field of gels, including molecular gels, in the intervening century [3, 20–25, 29]. They have been driven in part by advances in instrumentation which allows processes and structures to be interrogated at increasingly short time and length scales [26]. Despite these advances, some of the frustrations noted by Dorothy Jordan Lloyd nearly a century ago [27]—"The colloid condition, the gel, is one which is easier to recognize than to define"—persist today, and the field of molecular gels remains an active area of inquiry. Unfortunately, part of the reason for even greater progress is a lack of standardized, accepted methods for preparing and analyzing gels [28].

### 7.3 Molecular Gels and Approaches to Their Analyses

In this chapter, we focus on the structures and properties of molecular gels; the techniques for obtaining the structural information, usually indirect methods such as neutron and X-ray scattering, or direct methods such as atomic force and electron microscopies will not be discussed here. For an excellent coverage of methods to determine gel structures at different distance scales, see Ref. [29]. Only in a very limited number of examples have the molecular packing arrangements of gelator molecules within their gel networks been determined. The difficulties to obtain this information arise from several factors: (1) many molecular gelators are polymorphous, so that the phase within a gel may be different from that obtained from single crystal X-ray analyses, even if an appropriate crystal can be grown; (2) the gels networks are dynamic, so that the molecules within them are not fixed ‘permanently’ in space; (3) the habit within the crystalline network may differ depending on the liquid from which it is grown; (4) objects within the network may include disordered liquid molecules; (5) removal of the liquid from a gel risks changing the morph of the xerogel left behind; (6) intrinsically, the individual gelator objects within the gel network are very small, despite the gelator molecular sizes, shapes and conformational labilities varying over a wide range. An example of the latter is the family of acyclic to pentacyclic triterpenes whose structures and gelation properties have been compared [30]. Despite these difficulties, methods for determining the crystalline organization of gelator molecules within the gel networks are being developed and there is promise of others [31]. They rely on solid state (magic-angle) NMR techniques [32], synchrotron radiation analyses [33], and correlations between powder diffraction patterns from intrinsic gels and derived from single crystal data [34]. Also, the orientation of gelator molecules within a single gel fiber can be determined, in principle, from linearly polarized radiation and knowledge of the direction of the transition dipoles of the chromophores or lumophores [35]. Although this method holds great promise, it is still in a state of development and will not be applicable to all molecular gelator assemblies.

The structural analyses have been and continue to be aided by calculations at different levels of sophistication on single gelator molecules and ensembles of them. In that regard, density functional theory (DFT), molecular dynamics, statistical mechanical, and other types of calculations are being used to discern details of association between molecular gelators at the early and latter stages of aggregation, and to correlate the results of those calculations with experimental observations [36–39].

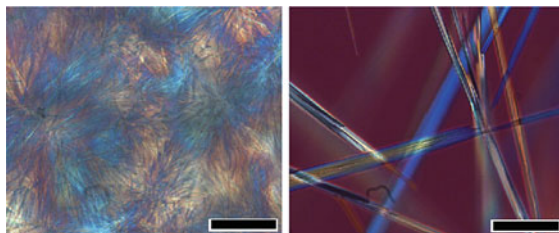
Even with these data, additional structural challenges will remain. By necessity, gelator networks must be 3-dimensional in order to immobilize the liquid component (usually by attractive short-range interfacial and capillary forces and longer-range repulsive interactions [23, 40]). Those networks rely on a fusion of the objects at ‘junction zones’ (i.e., at the intersection points between the constituent objects). Even though more information is forthcoming about the shapes and properties of the objects within the gelator networks, and they constitute the vast majority of the mass of the networks, very little is known about the nature of the molecular organizations

within zones, despite their playing a crucial role in determining the strength, elastic properties and longevity of the gels. Unfortunately, there does not appear to be a general means to assess the molecular packing within junction zones at this time, and it will be difficult to do so in the future because of their low concentrations and (presumably) greater disorder than within the gelator objects.

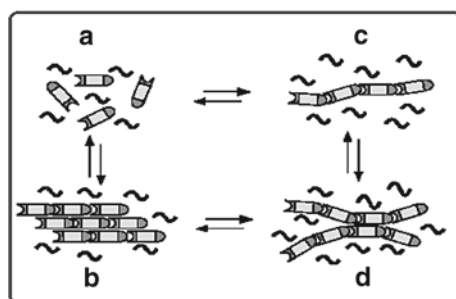
Their gel networks are comprised of molecules in which the networks responsible for providing the viscoelastic properties are linked physically through non-covalent interactions (usually, ionic, H-bonding and/or molecular dispersion forces). Many of them can be cycled with their corresponding sol/solution phases by heating above and cooling below their characteristic gelation temperatures; they are reversible thermally even if the non-covalent bonds are broken. Of course, there is a critical gelator concentration, usually defined at ambient temperature, below which percolation of the gelator molecules into fibrillar or other objects does not lead to a 3D network and gelation of the liquid component. Polymer gels, in which the networks responsible for providing the viscoelastic properties are from monomers held together by covalent bonds (i.e., linked chemically), will not be discussed except as needed to provide context for molecular gels. As opposed to many molecular gels (i.e., physical gels), polymer gels are not reversible thermally with their corresponding sol/solution phases if the covalent bonds of the polymer network are broken. In fact, Flory included molecular gels almost as an afterthought, presumably in his fourth class of gels [41]: (1) well-ordered lamellar structures; (2) cross-linked polymeric networks swollen with solvent; disordered polymer chains; (3) polymer networks in which the chain-chain interactions are physical; (4) particulate disordered structures.

## 7.4 Making Molecular Gels

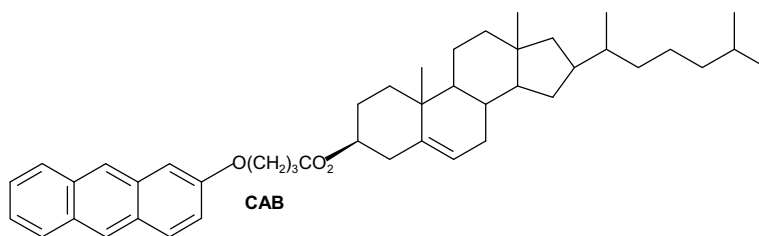
As mentioned above, the most common method to make molecular gels is by cooling their sols/solutions and by heating molecular gels to make sols/solutions. However, in each case, the shapes and sizes of the objects constituting the gel network depend on the rate of cooling of the sol phase (Fig. 7.2) [42] and critically on the detailed nature of solvent-gelator molecular interactions during the aggregation process leading to **D** in Fig. 7.3 [43]. The potential sensitivity of the gelator aggregation mode to cooling rate and liquid composition is clearly displayed by 1.5 wt% 3 $\beta$ -cholesteryl 4-(2-anthryloxy)butanoate (**CAB**) gels (Fig. 7.4) [44, 45]: they exhibit  $T_g$  values and fluorescence maxima at  $\sim 39^\circ\text{C}$  and 421 nm, respectively, in hexadecane and  $\sim 60^\circ\text{C}$  and 427 nm, respectively, in 1-octanol; the hexadecane-like or the 1-octanol-like gel network can be formed repeatedly and reproducibly when the sol phases are fast-cooled or slow-cooled and then reheated to the sol phase and re-cooled within a specific, intermediate range of liquid compositions. Neutron diffraction and X-ray scattering studies demonstrate that the packing arrangement of the **CAB** molecules and the shapes of the constituent fibers are different in the networks prepared by the different cooling protocols [46].



**Fig. 7.2** Polarized optical micrographs of gels at room temperature comprised of 2 wt% N,N'-dimethylurea in carbon tetrachloride that were formed by rapidly (spherulites; left) and very slowly (rods; right) cooling the sol phase. The scale bars are 200  $\mu\text{m}$ . Reprinted with permission from Chem Euro J 2005, 11, 3243. Copyright (2005) Wiley

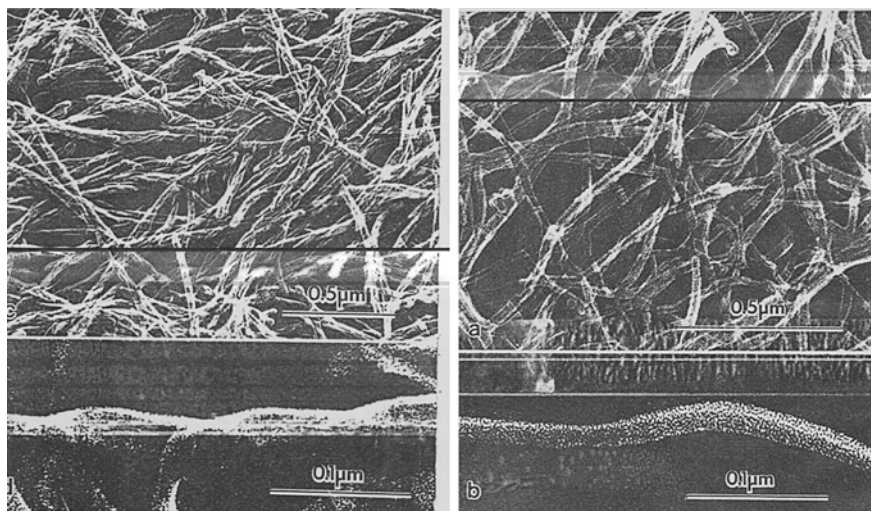


**Fig. 7.3** Cartoon representation of bullet-shaped molecules in a sol phase (a) aggregating to effect bulk crystallization (b) in competition with selective growth into a rod-like 1D structures (c) and then into 3D (d) networks. The liquid is represented by the squiggly-shaped lines. Adapted from Schoonbeek, Ph.D. Thesis, Univ. Groningen, The Netherlands, 2001



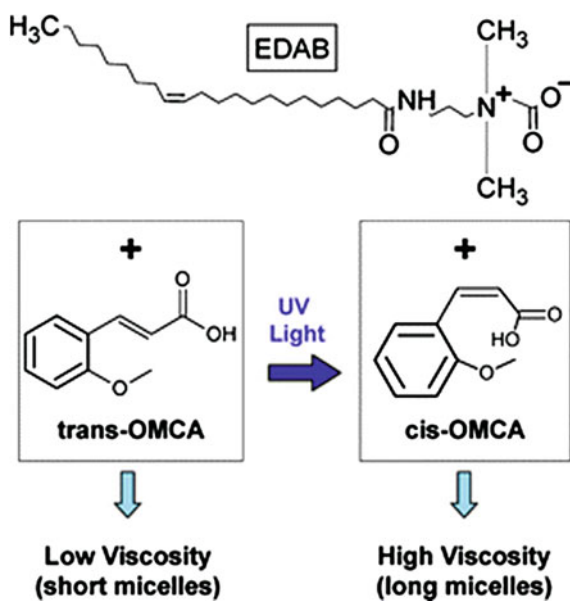
Other physical methods [47] include relieving or imposing a mechanical stress on thixotropic sols (vide infra), shining light on photoresponsive sols [48] (Fig. 7.5 [49]), effecting chemical changes to potential gelators by in situ enzymatic reactions [50], and placing sols under ultrasound to promote aggregation by conformational changes [51]. Thixotropic gels become fluid when disturbed (as by shaking) and then recover their viscoelasticity when left at rest. Some other chemical and physical methods to make or destroy molecular gels—some of which are reversible and some are irreversible—include changing the pH, adding or removing metal or other ions,





**Fig. 7.4** Freeze-fracture electron micrographs of CAB in 1-octanol (twisted fibers; left) and in n-dodecane (fibers with rectangular, untwisted fibers; right). Reprinted with permission from J Am Chem Soc 1989, 111, 5542. Copyright (1989) American Chemical Society

**Fig. 7.5** Influence of UV radiation on the formation of gels in sols composed of the surfactant EDAB to which some OMCA has been added. Reprinted with permission from Soft Matter 2009, 5, 797. Copyright (2009) Royal Society of Chemistry



initiating host-guest complexation, effecting redox processes, applying magnetic fields, and inducing in situ reversible or irreversible chemical reactions [52].

## 7.5 How Do Aggregation and Growth of Molecular Gelator Networks Occur?

The actual shapes of the micron-scale objects constituting the gelator network depend on the shape, chirality [53], and solubility of the gelator molecules [54] and kinetic factors (such as the rate of super-cooling) [10, 55], along the kinetic pathway leading to aggregates and their epitaxial growth into axially symmetric objects with large aspect ratios. The objects may have uniform or poly-disperse cross-sections as rods, straight or helically twisted tapes and fibers, tubules, spherulites, etc. In fact, helically twisted tapes can be produced even by achiral molecular gelators [56]! Several types of gelator networks, including those comprised of tetraoctadecylphosphonium salts, have been used as templates to make silicate objects [57] with tetraethyl orthosilicate (TEOS) precursors [58].

In fact, one can envision several mechanisms by which small molecules in sol phases aggregate into the 3D networks requisite for gel formation. A cartoon representation of one possible mode from a sol phase is shown in Fig. 7.3. The bullet-shaped molecules prefer to associate along one axis, leading to preferential 1D growth of rod-like objects shown in C. A macroscopic analogy is how Lego blocks interact preferentially along only their crenelated surfaces. Rather empirical treatments of the formation and structures of gel networks based on kinetic parameters have been devised by Avrami (Eq. 7.1) [59] and Dickinson (Eq. 7.2) [60]. The original articles describe the specific conditions for applying them to molecular gels.  $X$  is the volume fraction of the gelator participating in the gel network at time  $t$ ,  $K$  is a type of rate constant,  $n$  is the ‘Avrami exponent’ which characterizes the type of object growth,  $D_f$  is the fractal dimension of the gelator network, and  $C$  is a constant. The variable  $X$  has been measured as a function of time by absorption and fluorescence spectroscopies, as well as by rheological and small angle neutron diffraction measurements; any technique which measures rapidly the changes in gelator aggregation may be employed. Other, more detailed approaches to aggregation/nucleation/growth mechanisms are based on isodesmic and cooperative modes of aggregation [61] and a combination of kinetic and thermodynamic considerations [62].

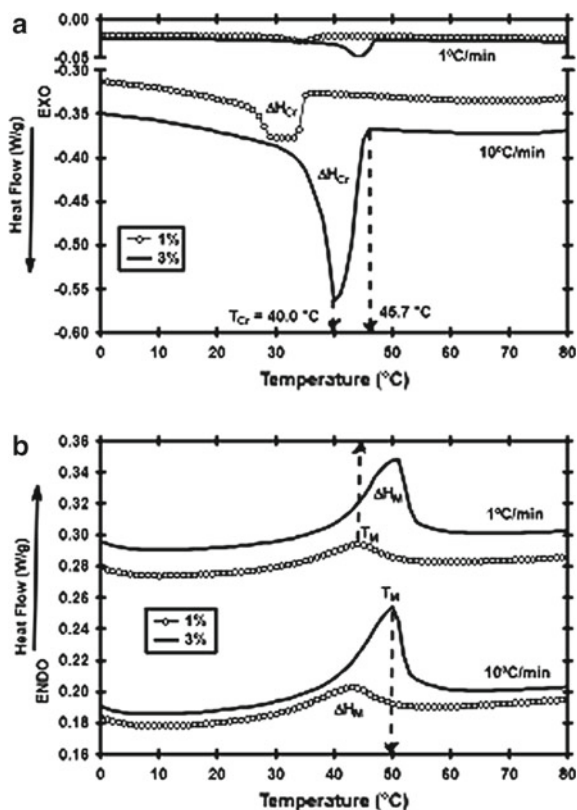
$$\ln[\ln(1 - X)^{-1}] = \ln K + n \ln t \quad (7.1)$$

$$\ln X = C + (3 - D_f)/D_f \ln t \quad (7.2)$$

## 7.6 Experimental Determination of What Is and Is Not a Gel

A simple and preliminary (but sometimes inaccurate) method to determine what is and what is not a gel is 'the inverse flow method' [63] in which a sample consisting of a liquid and a small fraction of solid is inverted with respect to gravity. If no flow is observed after a protracted period, the sample may be assumed to be a gel. A more accurate and detailed assessment of the gel status of a sample is available from measurements of endotherms and exotherms in differential scanning calorimetry thermograms. An example is the thermal behavior of 1 and 3 wt% *n*-dotriacontane ( $C_{32}H_{66}$ ) in safflower oil rich in trioleins [64]. From the thermograms obtained on cooling the sol (Fig. 7.6a) and heating the gel (Fig. 7.6b) at 1 or 10 °C, it was possible to calculate the initial crystallization and melting temperatures, the temperatures of maximum heat flow for crystallization/melting ( $T_{Cr}$  and  $T_M$ , respectively), and the heats of crystallization/melting ( $\Delta H_{Cr}$  and  $\Delta H_M$ , respectively). These data allow interesting and mechanistically informative comparisons between thermodynamic and rheological changes (the latter obtained in separate experiments). They

**Fig. 7.6** Cooling (a) and heating (b) thermograms at 1 or 10 °C/min of dotriacontane ( $C_{32}H_{66}$ ) in a triolein-rich safflower oil. The arrows in A indicate the initial ( $T_{Cr} = 45.7$  °C) crystallization and maximum heat flow ( $T_{Cr} = 40.0$  °C) temperatures. The heat of crystallization,  $\Delta H_{Cr}$ , is obtained from the area under the exotherm. In B, the arrows show the melting temperatures ( $T_M$ ), and the areas under the corresponding endotherms provide the heats of melting,  $\Delta H_M$ . Reprinted with permission from European Journal of Lipid Science and Technology, 2009, 111, 207. Copyright (2009) Wiley-VCH



also provide insights into the validity of applying the Schröder-van Laar equation [65] (Eq. 7.3), which assumes ideality in melting-nucleation phenomena during the sol-gel phase transitions (i.e., the interactions between gelator and liquid molecules is the same in the sol and gel phases) over a range of concentrations). In Eq. 7.3,  $x$  is the gelator solubility under conditions of ideal solution behavior,  $T$  is the equilibrium temperature,  $R$  is the ideal gas constant, and  $\Delta H_M$  and  $T_M$ —available from DSC measurements—are, respectively, the molar melting enthalpy and the melting temperature of the neat gelator. Because the assumption of ideality is frequently not valid, the use of this equation is useful in a limited number of gel systems.

$$\ln x = (\Delta H_M/R)(1/T_M - 1/T) \quad (7.3)$$

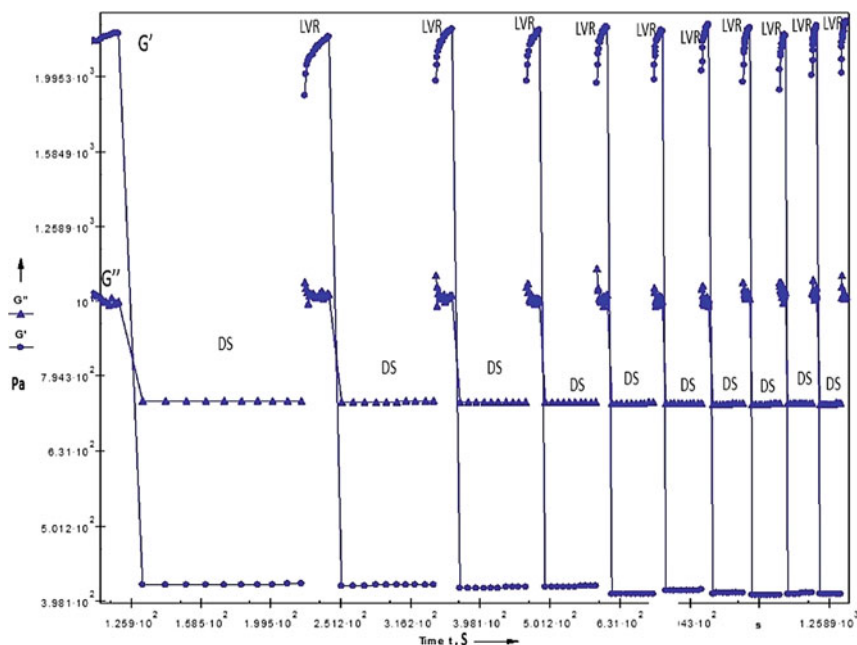
Thixotropic gels are especially interesting because they can be destroyed by application of excessive stress and reformed when the sample is left under conditions within the linear viscoelastic region, which is determined by oscillatory rheological measurements of the moduli as a function of oscillation frequency and applied stress. The values of the moduli, as a sample is cycled periodically between the linear viscoelastic (LVR) and the destructive strain (DS) regions, can yield valuable insights into the dynamics of the isothermal reformation step and the degree (if any) of irreversible destruction of the gelator structure upon repeated application of destructive strain. In the example shown in Fig. 7.7 with the gelator (*R*)-12-hydroxy-N-(2-hydroxyethyl)stearamide (**HS-2-OH**), there is no discernible loss of gel structure after repeated exposure to destructive strain [66]. That is usually not the case: the magnitudes of the moduli decrease as the number of cycles or the magnitude of the applied strain increases. The recovery time constant  $\tau$  after cessation of destructive strain can be calculated from Eq. 7.4 (where  $m$  is a constant that depends on the gel being examined).

$$\ln \left[ -\ln \frac{G'(\infty) - G'(t)}{G'(\infty) - G'(0)} \right] = m \ln t - m \ln \tau \quad (7.4)$$

## 7.7 The Role of the Liquid Component

Despite the gelator being a much smaller portion of a gel than the liquid, the properties of the latter are frequently given less attention. To discern how to characterize the properties of the liquid, it is necessary to consider the interactions between the two components at various stages during gel formation. Unfortunately, methods to assess gelator-liquid interactions in sol phases, for example, are not well developed. Thus, most approaches concentrate on whether a particular liquid is gelled by a specific gelator molecule and comparisons of the properties of the two.

Several of those approaches have been compared recently [67b]. Here, the one of Hansen will be discussed as it applies to gels [67], comparing it with the more



**Fig. 7.7**  $G'$  (■) and  $G''$  (▲) versus time and application of different strains and frequencies (LVR = linear viscoelastic region; DS = destructive strain region) of a 2.0 wt% HS-2-OH in isostearyl alcohol sample at 20 °C. Reprinted with permission from Soft Matter 2015, 11, 5010. Copyright (2015) Royal Society of Chemistry

traditional Hildebrand solubility parameter [Eq. 7.5, where  $\Delta E_i^v$  is the energy of vaporization and  $V_i$  is the molar volume; includes dispersion forces and polar interactions (including H-bonding)] that emphasizes enthalpy without considering equally entropy. Then, Eq. 7.6 applies for a mixture containing 2 different components (e.g., a solvent and a gelator) where is the volume of the mixture and  $\phi_i$  is the volume fraction of component  $i$ .

$$\delta_i = \left( \frac{\Delta E_i^v}{V_i} \right)^{1/2} \quad (7.5)$$

$$\Delta H_m = V \left( \phi_1 \left[ \frac{\Delta E_1^v}{V_1} \right]^{1/2} - \phi_2 \left[ \frac{\Delta E_2^v}{V_2} \right]^{1/2} \right) \quad (7.6)$$

Hansen solubility parameters (HSPs) consider specific intermolecular interactions as 3 separate components for the energy of vaporization as the cohesive energy. They are London (atomic) dispersion forces ( $E_d$ ), (molecular) permanent dipole-dipole, quadrupole-quadrupole, ion-ion etc. forces ( $E_p$ ), and (molecular) hydrogen bonding ( $E_h$ ) and  $E_{total} = E_d + E_p + E_h$ . Then,  $E_{total}/V$  (in  $\text{J}/\text{cm}^3 = \text{MPa}$ ) is a pressure and

the sum of squares of the individual HSP components (i.e., the Hildebrand solubility parameter) is  $\delta_{total}^2 = \delta_d^2 + \delta_p^2 + \delta_h^2$ . From this, one can calculate Hansen spheres by calculating  $\delta_d$ ,  $\delta_p$  and  $\delta_h$  as the center of a sphere  $R_{ij}$  (Eq. 7.7; where  $i$  is a solvent and  $j$  is a gelator) on a group summation basis, using data from Hansen's book [67] or empirically if a sufficient number of gelled and non-gelled liquids have been employed.

$$R_{ij} = \sqrt{4(\delta_{di} - \delta_{dj})^2 + (\delta_{pi} - \delta_{pj})^2 + (\delta_{hi} - \delta_{hj})^2} \quad (7.7)$$

An attractive alternative to the Hansen solubility parameters is Teas plots, another empirical relationship between liquid properties and gelator (or other species) solubilities, [67a, 68], from which triangles of solubility can be constructed [69, 70].

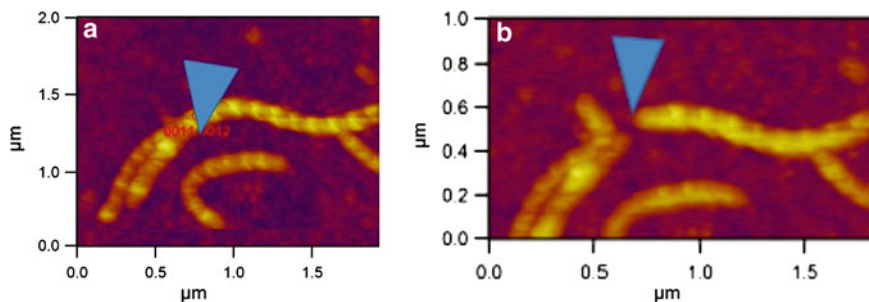
## 7.8 Organic Gelator Molecules and Their Assemblies. Starting from the Simplest Molecular Structures and Increasing the Complexity

A short synopsis of a limited number types of molecular gelators and their aggregate gel structures are presented below. The examples have been selected to give the reader a 'flavor' of the myriad of known molecular gels. Two of the most efficient molecular gelators, in terms of the range of liquids gelled and the CGCs ( $\sim .03$  to  $.05$  wt%) required, are methyl 4,6-O-*p*-nitrobenzylidene- $\alpha$ -D-galactopyranoside and methyl 4,6-O-benzylidene- $\alpha$ -D-mannopyranoside [71]. For a more comprehensive treatment of this topic, see the books and reviews cited.

### 7.8.1 *n*-Alkanes and Their Simple Derivatives

Very short *n*-alkanes melt to their isotropic liquid phases at sub-ambient temperatures and, therefore, are inappropriate candidates to form useful assemblies for gelation of other liquids. However, longer *n*-alkanes exhibit neat solid phases with more than one type of organization and super-ambient melting temperatures [72]. For example, *n*-heneicosane (C<sub>21</sub>H<sub>44</sub>) undergoes transitions from orthorhombic (Phase I) to hexagonal-rotator (Phase II) layers at 32.5 °C and from Phase II to its isotropic (liquid) phase at 40.2 °C [73]. Even in solid Phase II, due to thermal fluctuations, there is some disordering and *gauche* chain conformations which are less probable in a layer middle than at the layer ends [72]. In fact, very long *n*-alkanes are able to gelate short *n*-alkanes and a variety of liquids by forming networks consisting of 2D platelets [74].

The added disorder caused by inserting a carboxy group between carbons 4 and 5 of the *n*-heneicosane chain, yielding *n*-butyl stearate, induces formation of a smectic

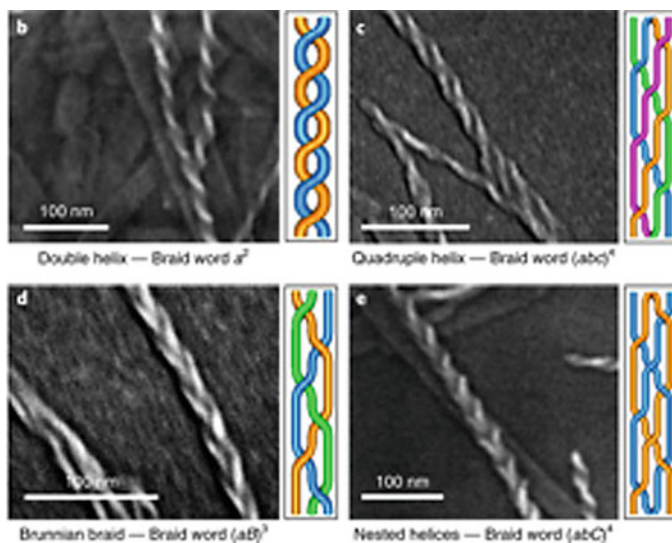


**Fig. 7.8** Atomic force microscope images of a 2 wt% **HS-2-OH** in silicone oil gel. The blue triangle is the tip before (a) and after (b) it was used to cleave one of the ropes. Reprinted with permission from *Soft Matter* 2015, 11, 5010. Copyright (2015) Royal Society of Chemistry

B liquid-crystalline phase between 15 and 26 °C [75]. Addition of a carboxylic acid group to a chain end of a long *n*-alkane reduces the disorder, and layered surfactant hydrogels are formed when water is added to the salts. For example, stearic acid (mp ~157 °C) is the carboxylic acid derivative of octadecane (mp 28 °C). Its potassium or rubidium salt or mixtures of potassium stearate and 1-octadecanol form well organized lamellar hydrogels [76].

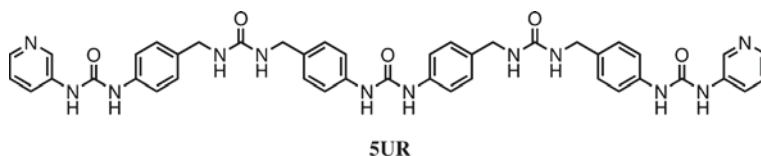
In fact, a large number of other end groups have been appended to long alkyl chains as a means to modify the melting temperature (i.e., making them well above room temperature) and solubility of potential gelators [77]. They, especially in combination with addition of a hydroxyl [78] or carbonyl group [79] near the chain middle (N. B.: at C12), have been shown to yield molecules capable of gelating a wide variety of liquids.

Whereas **HS-2-OH** [66] gelates several liquids by creating twisted ropes whose the pitch =  $130 \pm 30$  nm in silicone oil (Fig. 7.8), and  $60 \pm 5$  nm in isostearyl alcohol, [66] the shape of the objects responsible for the network of carbon tetrachloride gels supplied by the smallest molecular organogelator known to date, *N,N'*-dimethyl urea (MW 88), can be changed from spherulites (fast-cooling of sols) to rods (slow-cooling of sols) (Fig. 7.2) [42]. Even more surprising is the observation that the gel networks of achiral oligoureas can be various types of braided fibers. For example, the pentaurea **5UR** has been shown to form a variety of chiral braids in *N,N*-dimethylformamide (Fig. 7.9) as its sols are cooled to room temperature [80]. A detailed explanation for the different motifs has been reported, along with a caveat that an excess of one of the handedness within the braids can be attributed to traces of chiral species advertently (or inadvertently!) present in the sols. The gross aspects of their formation can be explained by a mathematical model reported more than 70 years ago [81]. Similar approaches have been employed to explain the change from fibers of sodium oleate/oleic acid in buffered aqueous media that are <100 mm-long helices to >1 cm long helical assemblies with a regular pitch and radius when small amounts of *N*-decanoyl-L-alanine are added to the initial solutions/sols [82].



**Fig. 7.9** SEM micrographs of braid topologies observed in gels of **5UR**. Reprinted with permission from Nature Chemistry 2019, 11, 375. Copyright (2019) Springer Nature

Less dramatic (but clear) elongation of the fibers was also found when some shorter saturated alkanolate salts were added in place of the *N*-decanoyl-L-alanine.

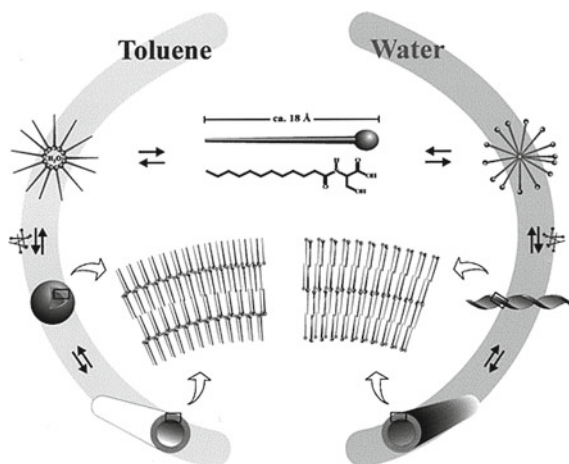


### 7.8.2 Some Tubule Assemblies of More Complex Alkane-Derived Surfactants

Nanotubules can offer an especially interesting mode of aggregation and growth leading to gel networks [53]. For example, the modes of growth and detailed structures of nanotubules from L-dodecanoylserine, a structurally simple molecule with a long alkyl chain and a polar head group, are very dependent on the liquid from which they are grown (Fig. 7.10) [83]. In water, aggregation proceeds from micelles and helical ribbons, leading eventually to tubules (mp 17–20 °C) in which the polar head groups pointed outward. When grown from toluene, the progression is from inverted micelles and vesicles and the lipophilic tails are pointed outward in the

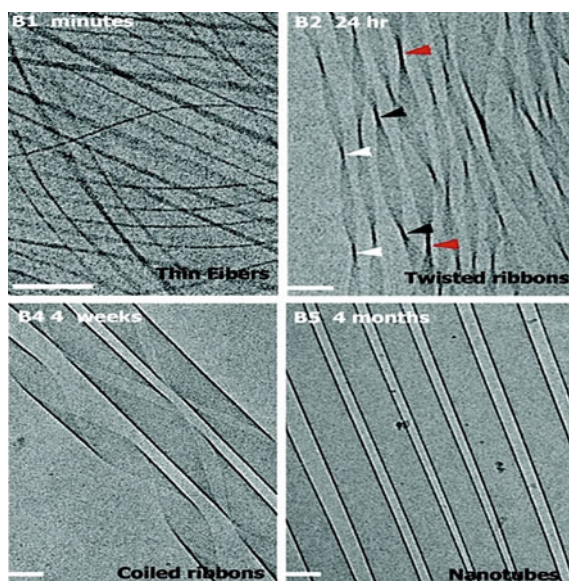


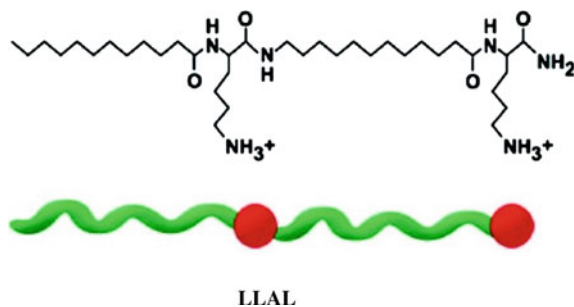
**Fig. 7.10** Progression of aggregation modes for L-dodecanoylserine into tubules grown from toluene and water. Reprinted with permission from Langmuir 2001, 17, 873. Copyright (2001) American Chemical Society



eventual tubules (mp 60–62 °C). Also, the ribbons are not twisted if the dodecanoylserine is racemic. Another detailed example uses cryo-TEM to follow the sequential transformation of aqueous dispersions of *N*- $\alpha$ -lauryl-lysyl-aminolauryl-lysyl-amide (**LLALA**), a molecule comprised of two dodecyl chains and two charged lysine groups, into 1D objects—from thin fibers/ribbons to twisted ribbons to coiled ribbons to closed nanotubes (Fig. 7.11) [84]. The observations of the separate objects are made possible by the very different rates for the transformation steps.

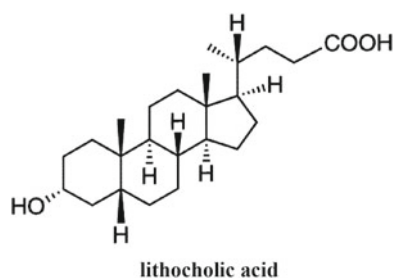
**Fig. 7.11** Temporal progression of LLALA to nanotubes (B5) by self-assembly in aqueous media at 25 °C: B1—thin micrometer-long fibrils; B2—twisted ribbons of various widths; B4—alternating arrowheads highlight a cylindrical curvature. Bars = 100 nm. Reprinted with permission from J Am Chem Soc 2011, 133, 2511. Copyright (2011) American Chemical Society



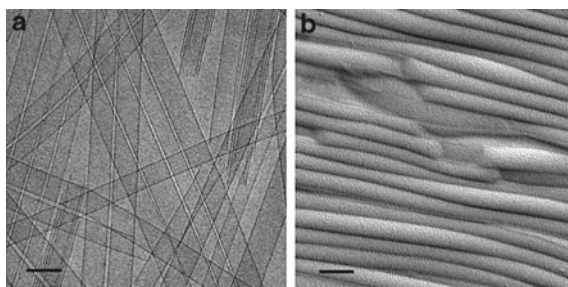


### 7.8.3 Some Tubule Assemblies of Steroidal Surfactants

Tubules with monodisperse diameters have been made by adding a simple bile acid, lithocholic acid, to an aqueous sodium hydroxide solution [85]. Single-walled tubules (52 and 49 nm outside and inside diameters) can be observed from the closure of twisted ribbons shortly after mixing (Fig. 7.12a). They are easily aligned and become multi-layered over longer times (Fig. 7.12b). They have been transformed into silica and titanium oxide objects with very large aspect ratios via sol-gel processes and calcination using tetraethyl orthosilicate (TEOS) and titanium(IV) isopropoxide ( $\text{Ti}(\text{OiPr})_4$ ), as the precursors [86].



**Fig. 7.12** Single-walled (**a**; cryo-TEM) and multi-walled tubules (**b**; freeze-fracture replicas) of sodium lithocholate that were formed in water. Bars = 100 nm. Reprinted with permission from Langmuir 2002, 18, 7240. Copyright (2002) American Chemical Society



## 7.9 Final Thoughts About the Scope and Intent of This Chapter

Unapologetically, this chapter relies more heavily on reports from research in the author's laboratory than a balanced rendering of the references warrants. That reliance is a result only of the author's familiarity and is not intended to be a balanced assessment of the importance of the contributions of others to the field.

With that in mind, this chapter is intended to describe what molecular gels are structurally and mechanically, and the known limits of both properties, and to describe the clear challenges to advancing the field; it is not intended to be a comprehensive treatise. It presents a description of what molecular gels are and how one can ascertain the difference them and other phases that depend on aggregation. It also describes some of the challenges confronting our ability to characterize them and, more importantly, how to design molecular gelators a priori and their interactions with the liquid components along the pathways starting from sols. Although molecular gels, comprised of ('0-D') gelator molecules, are intrinsically weaker mechanically than their polymeric gel analogues (in which at least one-dimension of the gel structure is always present through covalent intermolecular interactions among 'monomers'), they offer some potential advantages and opportunities. Foremost among these is their ease of reversibility that makes possible many interesting applications, some of which are described in several of the references cited here.

**Acknowledgements** RGW is extremely grateful to the many students in his laboratory and collaborators throughout the world who have helped to advance the field of this research and have been his educational guides. Much of the research from the lab at Georgetown would not have been possible without the financial support of the US National Science Foundation, most recently Grant CHE-1502856. This chapter is dedicated to the memory of two recently deceased outstanding scientists whose research in fields related to molecular gels has been inspirational to those directly in it. They are Kailasam Venkatesan of India (29 April 1932–31 December 2019) and Faruk Jose Nome Aguilera of Brazil who was born in Chile (29 May 1947–24 September 2018).

## References

1. Smith, D.K.: In: Weiss, R.G. (ed.) *Molecular Gels*. Royal Society of Chemistry, United Kingdom, chap 9 (2018)
2. (a) Hamley, I. W.: *Introduction to Soft Matter*. Wiley: Chichester. (b) Liu, X.Y., Li, J.L. (eds) *Soft Fibrillar Materials*. Wiley, Singapore (2013). (c) Doi, M.: *Soft Matter Physics*. Oxford University Press, Oxford (2013). (d) Miller, W.L., Cacciuto, A.: Exploiting classical nucleation theory for reverse self-assembly. *J. Chem. Phys.* **133**, 234108 (2010)
3. Marangoni, A.G., Garti, N. (eds.): *Edible Oleogels*, 2nd edn. AOCS Press, London (2018)
4. Ruiz-Palomo, C., Kennedy, S.R., Soriano, M.L., Jones, C.D., Valcarcel, M., Steed, J.W.: Pharmaceutical crystallization with nanocellulose organogels. *Chem. Commun.* **52**, 7782–7785 (2016)
5. Clark, J.B.: US Patent 2,596,844 13 May 1952
6. Fieser, L.F., Harris, G.C., Hershberg, E.B., Morgana, M., Novello, F.C.: Putnam ST Napalm. *Ind. Eng. Chem.* **38**, 768–773 (1946)

7. McNaught, A.D., Wilkinson, A.: IUPAC. Compendium of Chemical Terminology (the “Gold Book”), 2nd edn. Blackwell Scientific, Oxford (1997)
8. Dawn, A., Kumari, H.: Low Molecular Weight Supramolecular Gels Under Shear: Rheology as the Tool for Elucidating Structure-Function Correlation. *Chem. Eur. J.* **24**, 762–776 (2018)
9. Marangoni, A.G.: Kinetic Analysis of Food Systems. Springer, Switzerland, p. 161ff (2017)
10. Huang, X., Raghavan, S.R., Terech, P., Weiss, R.G.: Distinct Kinetic Pathways Generate Organogel Networks with Contrasting Fractality and Thixotropic Properties. *J. Am. Chem. Soc.* **128**, 15341–15352 (2006)
11. Fuentes-Caparrós, A.M., Dietrich, B., Thomson, L., Chauveau, C., Adams, D.J.: Using cavitation rheology to understand dipeptide-based low molecular weight gels. *Soft Matter* **15**, 6340–6347 (2019)
12. (a) Li, J.-J., Zhang, M., Weiss, R.G.: (R)-12-Hydroxystearic acid hydrazides as very efficient gelators: Diffusion, partial thixotropy, and self-healing in self-standing gels. *Chem. Asian J.* **11**, 3414–3422 (2016). (b) Song, J., Wang, H., Li, M.: An NMR study on the gelation of N,N'-bis(4-N-alkylo-xybenzoyl) hydrazine (4Dn) in two aromatic solvents. *New J. Chem.* **39**, 2711–2719 (2015)
13. Feng, L., Cavicchi, K.A.: Investigation of the relationships between the thermodynamic phase behavior and gelation behavior of a series of tripodal trisamide compounds. *Soft Matter* **8**, 6483–6492 (2012)
14. Takahashi, A., Sakai, M., Kato, T.: Melting temperature of thermally reversible gel. VI. Effect of branching on the sol–gel transition of polyethylene gels. *Polym. J.* **12**, 335–341 (1980)
15. Weiss, C.K., Toca-Herrera, J.L. (eds.): *Colloid Chemistry*. Basel, MDPI (2018)
16. Hsu, F.C., Tsai, S.F., Lee, S.-S.: Chemical investigation of Hyptis suaveolens seed, a potential antihyperuricemic nutraceutical, with assistance of HPLC-SPE-NMR. *J. Food Drug Anal.* **27**, 897–905 (2019)
17. (a) Dresel, W., Heckler, R.P.: Lubricating greases in lubricants and lubrication. In: Mang, T., Dresel, W. (eds) *Lubricants and Lubrication*. Wiley, Weinheim (2017). (b) Mortier, R.M., Fox, M.F., Orszulik, S.T.: *Chemistry and Technology of Lubricants*, 3rd edn. Springer, Dordrecht (2010)
18. Lipowitz, A.: Versuche und Resultate über die Löslichkeit der Harnsäure. *Liebigs Ann Chem Pharm* **38**, 348–355 (1841)
19. Graham, T.: X. liquid diffusion applied to analysis. *Phil. Trans. Roy Soc.* **151**, 183–224 (1861)
20. te Nijenhuis, K.: Thermoreversible networks. *Adv. Polym. Sci.* **130** (Springer Verlag, Berlin) (1997)
21. Escuder, B., Miravet, J.F. (eds.): *Functional Molecular Gelators*. RSC Publishing, Cambridge (2014)
22. Liu, X.Y., Li, J.L. (eds.): *Soft Fibrillar Materials: Fabrication and Applications*. Wiley-VCH Verlag, Weinheim (2013)
23. Weiss, R.G., Terech, P. (eds.): *Molecular Gels. Materials with Self-Assembled Fibrillar Networks*. Springer, Dordrecht (2006)
24. Weiss, R.G. (ed.): *Molecular Gels*. Royal Society of Chemistry, United Kingdom (2018)
25. Loh, X.J., Scherman, O.A. (eds.): *Polymeric and Self Assembled Hydrogels*. RSC Publishing, Cambridge (2013)
26. Weiss, R.G.: The past, present, and future of molecular gels. What is the status of the field and where is it going? *J. Am. Chem. Soc.* **136**, 7519–7530 (2014)
27. Lloyd, D.J.: The problem of gel structure. In: Alexander, J. (ed.) *Colloid Chemistry*, vol. 1, pp. 767–782. The Chemical Catalog Co, New York (1926)
28. Weiss, R.G.: Controlling variables in molecular gel science. How can we improve the state of the art? *Gels* **4**, 25ff (9 pages) (2018)
29. Guenet, J.M.: *Organogels. Thermodynamics, structure, solvent role, and properties*. Springer Nature, Switzerland (2016)
30. Bag, B.G., Majumdar, R.: Self-assembly of renewable nano-sized triterpenoids. *Chem. Rec.* **17**, 841–873 (2017)

31. Das, U.K., Banerjee, S., Dastidar, P.: Remarkable shape-sustaining, load-bearing, and self-healing properties displayed by a supramolecular gel derived from a bis-pyridyl-bis-amide of L-phenyl alanine. *Chem. Asian J.* **9**, 2475–2482 (2014)
32. Nonappa, Lahtinen M., Behera, B., Kolehmainen, E., Maitra, U.: Unraveling the packing pattern leading to gelation using SS NMR and X-ray diffraction: direct observation of the evolution of self-assembled fibers. *Soft Matter* **6**, 1748–1757 (2010)
33. Takeno, H., Mochizuki, T.: A structural development of an organogel explored by synchrotron time-resolved small-angle X-ray scattering. *Colloid Polym. Sci.* **291**, 2783–2789 (2013)
34. Ostuni, E., Kamaras, P., Weiss, R.G.: Novel X-ray method for *in situ* determination of Gelator strand structure. Cholesteryl Anthraquinone-2-carboxylate. *Angew. Chem. Int. Ed. Engl.* **35**, 1324–1326 (1996)
35. Giansante, C., Raffy, G., Schafer, C., Rahma, H., Kao, M.T., Olive, A.G.L., Del Guerzo, A.: White-light-emitting self-assembled nano fibers and their evidence by microspectroscopy of individual objects. *J. Am. Chem. Soc.* **133**, 316–325 (2011)
36. Dou, C.D., Li, D., Gao, H.Z., Wang, C.Y., Zhang, H.Y., Wang, Y.: Sonication-induced molecular gels based on mono-cholesterol substituted Quinacridone derivatives. *Langmuir* **26**, 2113–2118 (2010)
37. Vujicic, N.S., Glasovac, Z., Zweep, N., van Esch, J.H., Vinkovic, M., Popovic, J., Zinic, M.: Chiral hexa- and nonamethylene-bridged bis(L-Leu-oxalamide) Gelators: The first oxalamide gels containing aggregates with a chiral morphology. *Chem. Eur. J.* **19**, 8558–8572 (2013)
38. (a) Gránásy, L., Pusztai, T., Borzsonyi, T., Warren, J.A., Douglas, J.F.: A general mechanism of polycrystalline growth. *Nat. Mater.* **3**, 645–650 (2004). (b) Raghavan, S.R., Douglas, J.F.: The conundrum of gel formation by molecular nanofibers, wormlike micelles, and filamentous proteins: gelation without cross-links? *Soft Matter*. **8**, 8539–8546 (2012)
39. Sanz, E., White, K.A., Clegg, P.S., Cates, M.E.: Colloidal gels assembled via a temporary interfacial Scaffold. *Phys. Rev. Lett.* **103**, 255502 (2009)
40. Terech, P., Weiss, R.G.: Low-molecular mass gelators of organic liquids and the properties of their gels. *Chem. Rev.* **97**, 3133–3159 (1997)
41. Flory, P.: Thermodynamics of polymer solutions. *Disc Faraday Soc.* **49**, 7–29 (1970)
42. George, M., Tan, G., John, V.T., Weiss, R.G.: Urea and Thiourea derivatives as low molecular-mass organogelators. *Chem. Euro J.* **11**, 3243–3254 (2005)
43. (a) Lan, Y., Corradini, M.G., Weiss, R.G., Raghavan, S.R., Rogers, M.A.: To gel or not to gel: Correlating molecular gelation with solvent parameters. *Chem. Soc. Revs.* **44**, 6035–6058. (b) Hirst, A.R., Coates, I.A., Boucheteau, T.R., Miravet, J.F., Escuder, B., Castelletto, V., Hamley, I.W., Smith, D.K.: Low-molecular-weight gelators: elucidating the principles of gelation based on gelator solubility and a cooperative self-assembly model. *J. Am. Chem. Soc.* **130**, 9113–9121 (2008). (c) Edwards, W., Lagadec, C.A., Smith, D.K.: Solvent-gelator interactions-using empirical solvent parameters to better understand the self-assembly of gel-phase materials. *Soft Matter*. **7**, 110–117 (2011). (d) Chen, J., Kampf, J.W., McNeil, A.J.: Comparing molecular gelators and nongelators based on solubilities and solid-state interactions. *Langmuir* **26**, 13076–13080 (2010) (e) Muro-Small, M.L., Chen, J., McNeil, A.J.: Dissolution parameters reveal role of structure and solvent in molecular gelation. *Langmuir* **27**, 13248–13253 (2011). (f) Jonkheijm, P., van der Schoot, P., Schenning, A.P.H.J., Meijer, E.W.: Probing the solvent-assisted nucleation pathway in chemical self-assembly. *Science* **313**, 80–83 (2006). (g) De Greef, T.F.A., Smulders, M.M.J., Wolfs, M., Schenning, A.P.H.J., Sijbesma, R.P., Meijer, E.W.: Supramolecular polymerization. *Chem. Rev.* **109**, 5687–5754 (2009). (h) Aggeli, A., Nyrkova, I.A., Bell, M., Harding, R., Carrick, L., McLeish, T.C.B., Semenov, A.N., Boden, N.: Hierarchical self-assembly of chiral rod-like molecules as a model for peptide beta-sheet tapes, ribbons, fibrils, and fibers. *Proc. Natl. Acad. Sci. USA* **98**, 11857–11862 (2001). (i) Van derHart, D.L., Douglas, J.F., Hudson, S.D., Antonucci, J.M., Wilder, E.A.: NMR characterization of the formation kinetics and structure of di-o-benzylidene sorbitol gels self-assembled in organic solvents. *Langmuir* **27**, 1745–1757 (2011). (j) De Yoreo, J.J., Vekilov, P.G.: Principles of crystal nucleation and growth. In: Dove, P.M., De Yoreo, J.J., Weiner, S. (eds.) *Biom mineralization*

- (Reviews in Mineralogy and Geochemistry). Mineralogical Society of America, Washington DC **54**, 57–93 (2003)
44. (a) Lin, Yc., Weiss, R.G.: A novel gelator of organic liquids and the properties of its gels. *Macromolecules* **20**, 414–417 (1987). (b) Lin, Yc., Kachar, B., Weiss, R.G.: A novel family of Gelators of organic fluids and the structure of their gels. *J. Am. Chem. Soc.* **111**, 5542–5551 (1989)
  45. Furman, I., Weiss, R.G.: Factors influencing the formation of thermally reversible gels comprised of cholesteryl 4-(2-Anthryloxy)butanoate in hexadecane, 1-octanol, or their mixtures. *Langmuir* **9**, 2084–2088 (1993)
  46. Terech, P., Furman, I., Weiss, R.G.: Structures of organogels based upon cholesteryl 4-(2-anthryloxy)butanoate, a highly efficient luminescing gelator: Neutron and X-ray small-angle scattering investigations. *J. Phys. Chem.* **99**, 9558–9566 (1995)
  47. Caran, K.L., Lee, D.C., Weiss, R.G.: Molecular gels and their fibrillar networks In: Liu, X.Y., Li J.L. (eds.) *Soft Fibrillar Materials: Fabrication and Applications*. Wiley-VCH Verlag, Weinheim, chap 1 (2013)
  48. Murata, K., Aoki, M., Suzuki, T., Harada, T., Kawabata, H., Komori, T., Ohseto, F., Ueda, K., Shinkai, S.: Thermal and light control of the sol-gel phase transition in cholesterol-based organic gels. novel helical aggregation modes as detected by circular dichroism and electron microscopic observation. *J. Am. Chem. Soc.* **116**, 6664–6676 (1994)
  49. Kumar, R., Raghavan, S.R.: Photogelling fluids based on light-activated growth of zwitterionic wormlike micelles. *Soft Matter* **5**, 797–803 (2009)
  50. (a) Zhang, Y., Kuang, Y., Gao, Y., Xu, B.: Versatile small-molecule motifs for self-assembly in water and the formation of biofunctional supramolecular hydrogels. *Langmuir* **27**, 529–537 (2011). (b) Du, X., Zhou, J., Xu, B.: Supramolecular hydrogels made of basic biological building blocks *chem Asian J.* **9**, 1446–1472 (2014)
  51. (a) Caran, K.L., Lee, D.C., Weiss, R.G.: Molecular gels and their fibrillar networks. In: Liu, X.Y., Li, J.L. (eds.) *Soft Fibrillar Materials: Fabrication and Applications*. Wiley-VCH Verlag, Weinheim chap 1.32, p. 40ff (2013). (b) Liu, J., He, P., Yan, J., Fang, X., Peng, J., Liu, K., Fang, Y.: An organometallic super-gelator with multiple-stimulus responsive properties. *Adv. Mater.* **20**, 2509–2511 (2008). (c) Naota, T., Koori, H.: Molecules that assemble by sound: An application to the instant gelation of stable organic fluids. *J. Am. Chem. Soc.* **127**, 9324–9325 (2005). (d) Wang, C., Zhang, D., Zhu, D.: A low-molecular-mass Gelator with an electroactive Tetrathiafulvalene group: Tuning the gel formation by charge-transfer interaction and oxidation. *J. Am. Chem. Soc.* **127**, 16372–16373 (2005). (e) Bardeling, D.: *Ultrasound Induced Gelation: A Paradigm Shift*, vol. 5, pp. 1969–1971 (2009). (f) Ogata, K., Naota, T., Takaya, H.: Metal array fabrication based on ultrasound-induced self-assembly of metalated dipeptides. *Dalton Trans.* **42**, 15,953–15,966 (2013)
  52. Mishra, R.K., Das, S., Vedhanarayanan, B., Das, G., Praveen, V.K., Ajayaghosh, A.: Stimuli-responsive supramolecular gels In: Weiss, R.G. (ed.) *Molecular Gels*. Royal Society of Chemistry, United Kingdom, chap 7 (2018)
  53. (a) Selinger, J.V., Spector, M.S., Schnur, J.M.: Theory of self-assembled tubules and helical ribbons. *J. Phys. Chem. B* **105**, 7157–7169 (2001). (b) ten Eikelder H.M.M., Huub, M.M., Markvoort, A.J., de Greef, T.F.A., Hilbers, P.A.J.: An equilibrium model for chiral amplification in supramolecular polymers. *J. Phys. Chem. B* **116**, 5291–5301
  54. (a) Yu, R., Lin, N.B., Yu, W.D., Liu, X.Y.: Crystal networks in supramolecular gels: formation kinetics and mesoscopic engineering principles. *CRYSTENGGCOMM* **17**, 7986–8010 (2015). (b) Yan, N., Xu, Z.Y., Diehn, K.K., Raghavan, S.R., Fang, Y., Weiss, R.G.: Correlations of gel properties with gelator structures and characterization of solvent effects. *Langmuir* **29**, 793–805 (2013). (c) Yan, N., Xu, Z., Diehn, K.K., Raghavan, S.R., Fang, Y., Weiss, R.G.: How Do Liquid Mixtures Solubilize Insoluble Gelators? Self-Assembly Properties of Pyrenyl-Linker-Glucono Gelators in Tetrahydrofuran-Water Mixtures. *J. Am. Chem. Soc.* **135**, 8989–8999 (2013). (d) Liu, S.J., Yu, W., Zhou, C.: Solvents effects in the formation and viscoelasticity of DBS organogels. *Soft Matter.* **9**, 864–874 (2013)

55. (a) Huang, X., Terech, P., Raghavan, S.R., Weiss, R.G. (2005) Kinetics and structure during 5 $\alpha$ -Cholestan-3 $\beta$ -yl *N*-(2-Naphthyl)carbamate/*n*-alkane organogel formation. *J. Am. Chem. Soc.* **127**, 4336–4344. (b) Li, J.L., Yuan, B., Liu, X.Y., Wang, R.Y., Wang, X.G.: Control of crystallization in supramolecular soft materials engineering. *Soft Matter*. **9**, 435–442 (2013). (c) Filobelo, L.F., Galkin, O., Vekilov, P.G.: *J. Chem. Phys.* **123**, 014904 (2005)
56. Liu, Y., Jia, Y., Zhu, E., Liu, K., Qiao, Y., Che, G., Yin, B.: Supramolecular helical nanofibers formed by an achiral monopyrrolotetrahydrofulvalene derivative: water-triggered gelation and chiral evolution. *New J. Chem.* **41**, 11060–11068 (2017)
57. (a) Jung, J.H., Ono, Y., Shinkai, S.: Sol-gel polycondensation of tetraethoxysilane in a cholesterol-based organogel system results in chiral spiral silica. *Angew. Chem. Int. Ed.* **39**, 1862–1865 (2000). (b) Jung, J.H., Ono, Y., Sakurai, K., Sano, M., Shinkai, S.: Novel vesicular aggregates of crown-appended cholesterol derivatives which act as gelators of organic solvents and as templates for silica transcription. *J. Am. Chem. Soc.* **122**, 8648–8653 (2000)
58. Huang, X., Weiss, R.G.: Silica structures templated on fibers of tetraalkylphosphonium salt gelators in organogels. *Langmuir* **22**, 8542–8552 (2006)
59. (a) Avrami, M.: Kinetics of phase change. I General Theory. *J. Chem. Phys.* **7**, 1103–1112 (1939). (b) Avrami, M.: Kinetics of phase change. II Transformation-time relations for random distribution of nuclei. *J. Chem. Phys.* **8**, 212–224 (1940)
60. Dickinson, E.: On gelation kinetics in a system of particles with both weak and strong interactions. *J. Chem. Soc., Faraday Trans.* **93**, 111–114 (1997)
61. Smulders, M.M.J., Nieuwenhuizen, M.M.L., de Greef, T.F.A., van der Schoot, P., Schenning, A.P.H.J., Meijer, E.W.: How to distinguish isodesmic from cooperative supramolecular polymerisation. *Chem. Eur. J.* **16**, 362–367 (2010)
62. Douglas, J.F., Dudowicz, J., Freed, K.F.: Lattice model of equilibrium polymerization. VII. Understanding the role of “cooperativity” in self-assembly. *J. Chem. Phys.* **128**, 224901 (2008)
63. Raghavan, S.R., Cipriano, B.H.: Gel formation: Phase diagrams using tabletop rheology and calorimetry. In: Weiss, R.G., Terech, P. (eds.) *Molecular Gels. Materials with Self-Assembled Fibrillar Networks*. Kluwer Academic Publishers, Dordrecht chap 8 (2006)
64. Morales-Rueda, J.A., Dibildox-Alvarado, E., Charó-Alonso, M.A., Weiss, R.G., Toro-Vazquez, J.F.: Thermo-mechanical properties of safflower oil organogels with candelilla wax or hentriacontane as gelator. *Eur. J. Lipid Sci. Tech.* **111**, 207–215 (2009)
65. (a) Beckman, W.: Crystallization: Basic Concepts and Industrial Applications, pp. 37–74. Wiley, Weinheim. (2013). (b) Feng, L. Ph.D. thesis. Investigation of the relationships between the thermodynamic phase behavior and gelation behavior of a series of tripodal trisamide compounds. University of Akron, pp. 13–16 (2012)
66. Mallia, V.A., Weiss, R.G.: Structural bases for mechano-responsive properties in molecular gels of (R)-12-hydroxy-N-( $\omega$ -hydroxyalkyl)octadecanamides. Rates of formation and responses to destructive strain. *Soft Matter*. **11**, 5010–5022 (Correction: 2015, 11, 5168) (2015)
67. (a) Hansen, C.M.: Hansen Solubility Parameters, 2nd edn. CRC Press, Boca Raton. (2007). (b) Lan, Y., Corradini, M.G., Weiss, R.G., Raghavan, S.R., Rogers, M.A.: To gel or not to gel: Correlating molecular gelation with solvent parameters. *Chem. Soc. Revs.* **44**, 6035–6058 (2015). (c) Singh, A., Auzanneau, F.I., Corradini, M.G., Grover, G., Weiss, R.G., Rogers, M.A.: Molecular nuances governing the self-assembly of 1,3:2,4-dibenzylidene-D-sorbitol. *Langmuir* **33**, 10907–10916 (2017). (d) Xu, H.Q., Song, J., Tian, T., Feng, R.X.: Estimation of organogel formation and influence of solvent viscosity and molecular size on gel properties and aggregate structures *Soft Matter*. **8**, 3478–3486 (2012)
68. Zhao, C.X., Wang, H.T., Li, M.: Research progress in the correlation between gelation properties and solvent parameters. *Acta Phys-Chim Sinica* **30**, 2197–2209 (2014)
69. Barton, A.F.M.: Handbook of solubility parameters and other cohesive parameters. CRC Press, Boca Raton, FL, 2nd edn (1991)
70. Vila, A.S., Garcia, J.M.B.: Extended abstract—computer applications and cleaning: teas fractional solubility parameter system. *Conserv. Smithsonian. Contrib. Conserv. Sci.* 35–38 (2012). <https://pdfs.semanticscholar.org/610c/be33953b4b1b88a31ff259cf9c897aacbb1c.pdf>

71. (a) Luboradzki, R., Gronwald, O., Ikeda, A., Shinkai, S.: Sugar-integrated “Supergelators” which can form organogels with 0.03–0.05% [g mL<sup>-1</sup>]. *Chem Lett.* **29**, 1148–1149 (2000). (b) Gronwald, O., Shinkai, S.: Sugar-integrated gelators of organic solvents. *Chem Eur. J.* **7**, 4328–4334 (2001)
72. Maroncelli, M., Strauss, H.L., Snyder, R.G.: The distribution of conformational disorder in the high-temperature phases of the crystalline *n*-alkanes. *J. Chem. Phys.* **82**, 2811–2824 (1985)
73. Schaefer, A.A., Busso, C.J., Smith, A.E., Skinner, L.B.: Properties of pure normal alkanes in the C<sub>17</sub> to C<sub>36</sub> Range. *J. Am. Chem. Soc.* **77**, 2017–2019 (1955)
74. Abdallah, D.J., Weiss, R.G.: *n*-Alkanes Gel *n*-Alkanes (and Many Other Organic Liquids). *Langmuir* **16**, 352–355 (2000)
75. Krishnamurthi, D., Krishnamourthy, K.S., Shashidar, R.: Thermal, optical, X-ray, infrared and NMR studies on the *a*-phase of some saturated aliphatic esters. *Mol. Cryst. Liq. Cryst.* **8**, 339–366 (1969)
76. (a) Vincent, J.M., Skoulios, A.: ‘Gel’ et ‘coagel’. I. Identification. Localisation dans un diagramme de phases et détermination de la structure du ‘gel’ dans le cas du stéarate de potassium. *Acta Cryst.* **20**, 432–441 (1966). (b) Vincent, J.M., Skoulios, A.: ‘Gel’ et ‘coagel’. II. Etude comparative de quelques amphiphiles *Acta Cryst.* **20**, 441–447 (1966). (c) Vincent, J.M., Skoulios, A.: ‘Gel’ and ‘coagel’. III. ‘Gel’ investigation in the equimolecular mixture potassium stearate-*n*-octadecanol. *Acta Cryst.* **20**, 447–451 (1966)
77. George, M., Weiss, R.G.: Molecular organogels. Soft matter comprised of low molecular-mass organic gelators and organic liquids. *Acc. Chem. Res.* **39**, 489–497 (2006)
78. (a) Mallia, V.A., George, M., Blair, D.L., Weiss, R.G.: Robust organogels from nitrogen-containing derivatives of (*R*)-12-hydroxystearic acid as gelators. Comparisons with gels from stearic acid derivatives. *Langmuir* **25**, 8615–8625 (2009). (b) Mallia, V.A., Terech, P., Weiss, R.G.: Correlations of properties and structures at different length scales of hydro- and organogels based on *N*-alkyl-(*R*)-12-hydroxyoctadecylammonium chlorides. *J Phys. Chem. B* **115**, 12401–12414 (2011). (c) Mallia, V.A., Weiss, R.G.: Self-assembled fibrillar networks and molecular gels employing 12-hydroxystearic acid and its isomers and derivatives. *J. Phys. Org. Chem.* **27**, 310–315 (2014). (d) Abraham, S., Lan, Y., Lam, R.S.H., Grahame, D.A.S., Kim, J.J.H., Weiss, R.G., Rogers, M.A.: Influence of positional isomers on the macroscale and nanoscale of aggregates of racemic hydroxyoctadecanoic acids in their molecular gel, dispersion and solid states. *Langmuir* **28**, 4955–4964 (2012). (e) Zhang, M., Selvakumar, S., Zhang, X., Sibi, M.P., Weiss, R.G.: Structural and solubility parameter correlations of gelation abilities for dihydroxylated derivatives of long-chained, naturally-occurring fatty acids. *Chem. Eur. J.* **21**, 8530–8543 (2015)
79. (a) Pal, A., Abraham, S., Roger, M.A., Dey, J., Weiss, R.G.: Comparison of dipolar, h-bonding, and dispersive interactions on gelation efficiency of positional isomers of keto and hydroxy substituted octadecanoic acids. *Langmuir* **29**, 6467–6475 (2013). (b) Zhang, M., Weiss, R.G.: Mechano-responsive, thermo-reversible, luminescent organogels derived from a long-chained, naturally-occurring fatty acid. *Chem. Eur. J.* **22**, 8262–8272 (2016). (c) Zhang, M., Weiss, R.G.: Mechano-switchable, luminescent gels derived from salts of a long-chained, fatty-acid gelator. *Phys. Chem. Chem. Phys.* **18**, 20399–20409 (2016)
80. Jones, C.D., Simmons, H.T.D., Horner, K.E., Liu, K.Q., Thompson, R.L., Steed, J.W.: Braiding, branching and chiral amplification of nanofibres in supramolecular gels. *Nat. Chem.* **11**, 375–381 (2019)
81. Artin, E.: Theory of braids. *Ann. Math.* **48**, 101–126 (1947)
82. Kageyama, Y., Ikegami, T., Hiramatsu, N., Takeda, S., Sugawara, T.: Structure and growth behavior of centimeter-sized helical oleate assemblies formed with assistance of medium-length carboxylic acids. *Soft Matter* **11**, 3550–3558 (2015)
83. Boettcher, C., Schade, B., Fuhrhop, J.H.: Comparative cryo-electron microscopy of noncovalent *N*-dodecanoyl- (D- and L-) serine assemblies in vitreous toluene and water. *Langmuir* **17**, 873–877 (2001)
84. Ziserman, L., Lee, H.Y., Raghavan, M., Danino, D.: Unraveling the mechanism of nanotube formation by chiral self-assembly of amphiphiles. *J. Am. Chem. Soc.* **133**, 2511–2517 (2011)



85. (a) Terech, P., de Geyer, A., Struth, B., Talmon, Y.: Self-assembled monodisperse steroid nanotubes in water. *Adv. Mater.* **14**, 496–498 (2002). (b) Terech, P., Talmon, Y.: Aqueous suspensions of steroid nanotubules: Structural and rheological characterizations. *Langmuir* **18**, 7240–7244 (2002)
86. Huang, X., Weiss, R.G.: Rodlike silica and titania objects templated on extremely dilute aqueous dispersions of self-assembled sodium lithocholate nanotubes. *J. Coll. Interface Sci.* **313**, 711–716 (2007)

# Chapter 8

## Fréedericksz-Like Positional Transition Triggered by An External Electric Field



Ke Xiao and Chen-Xu Wu 

**Abstract** Microparticles (colloidal particles) of different shapes suspended in an anisotropic nematic liquid crystal (NLC) host medium are important soft matter systems which behave quite differently from those simply composed of microparticles or conventional isotropic liquids. The embedded microparticles disturb the alignment of LC molecules and induce elastic distortions, generating long-range anisotropic interactions and topological defects. The replacement of isotropic liquids with NLC medium gives rise to abundant physical behaviors of the microparticles, which leads to a broad range of practical applications ranging from biological detectors to new display and topological memory devices. This chapter is devoted to the new dynamic behaviors of a microparticle suspended in a uniform nematic liquid crystal (NLC) cell in the presence of an external electric field, an important tool in soft matter systems to manipulate microparticles. Investigating the basic dependence of critical electric value on cell thickness, Frank elastic constant, microparticle size and density is essential for understanding the dynamical behaviors of microparticles. This chapter is organized as follows. We start with a short introduction on liquid crystal together with a review of the related literature on microparticle-suspended liquid crystal (Sect. 8.1). The theoretical background of liquid crystals with particular focus on order parameter, Frank-Oseen free energy, surface anchoring free energy, Fréedericksz transition, and multipole expansion is described in Sect. 8.2. The main theoretical model and tools are then outlined in Sect. 8.3 to study the properties of a single particle in a uniform nematic liquid crystal cell in the presence of an external electric field. The main results and discussions based on the theoretical model we proposed in Sect. 8.3 are presented in Sect. 8.4. In Sect. 8.5 a brief summary is made.

**Keywords** Nematic liquid crystal · Positional transition · Effective elastic energy · Green's function method

---

K. Xiao · C.-X. Wu (✉)

Department of Physics, College of Physical Science and Technology,  
Xiamen University, Fujian, People's Republic of China  
e-mail: [cxwu@xmu.edu.cn](mailto:cxwu@xmu.edu.cn)

## 8.1 Introduction

Liquid crystals (LCs) are soft materials made of organic molecules with rodlike, disclike or banana shapes, a mesophase intermediate between the crystalline and isotropic liquid state formed at a certain temperature or molecular concentration range [1, 2]. In the undeformed ground state, uniaxial molecules in nematic liquid crystal (NLC) phase, the simplest type of all orientational orders, prefer an orientation with their molecular long axes aligning along a common direction  $\mathbf{n}$  called director. It is widely accepted that the NLCs possess many anisotropic physical properties that are easy to control by external stimuli owing to their long-range orientational order rather than translational order. Colloids, which are widely used in our daily life, including milk, ink, paint, cream and fog, are dispersions of solid, liquid or gas particles with typical size ranging from a few nanometers up to a few micrometers in a host surrounding medium [3, 4]. When dispersed in a NLC, the colloidal particles disturb the alignment of LC molecules and induce elastic distortions which give rise to long-range anisotropic interactions and topological defects. The generated long-range force leads to the self-assembly of molecules in such system, a topological phenomenon offering the possibility to control and design anticipated function for novel composite materials and diverse topology materials with similar features. One of the main themes of liquid crystal is to study the properties and behaviors of colloids suspended in nematic liquid crystal (NLC), and a wide range of promising practical applications have been realized, such as new display and topological memory devices [5–7], new materials [8], external triggers and release microcargo [9], and biological detectors [10, 11]. Over the past two decades, many experimental, theoretical and computer simulation studies have been focused on the physical properties of colloidal particles embedded within NLCs [12–25].

At the experimental level, diverse methods and techniques have been developed to measure the interaction force between particles in NLC in a direct manner [12, 26–29]. It has been found that the interaction force of spherical particles suspended in NLC is associated not only with interparticle distance and geometrical confinement [13], but also with the shape of particles which plays a crucial role in pair interaction and aggregation behaviors [17]. Whereas, in the presence of the electric field, fruitful fascinating physical phenomena such as levitation, lift, bidirectional motion, aggregation, Electrokinetic and superdiffusion [24, 30–32] have been observed for colloids dispersed in NLCs. On the other hand, theoretical modeling and computer simulation as useful complements to experiments, such as Landau-de Gennes (LdG) theory and elastic free energy method, have been carried out to interpret the nature of colloidal particles dispersed in NLCs. Generally Monte Carlo simulation [23, 33], lattice Boltzmann method [34, 35] and finite element method [13, 36–39] are common adopted techniques to minimize the LdG free energy functional. Except for the methods mentioned above, recently S. B. Chernyshuk and coauthors studied the interaction between colloidal particles in NLCs with or without external field by using Green's function method, and obtained general formulae for interaction energy between colloidal particles [40–42]. In the liquid crystal and particles coex-

istence system, it is observed that external field is able to drive particles apart [12], cause rotation [18] and alignment [43] of LC molecules, and even manipulate the equilibrium position of microdroplet [44]. Although interactions of two particles in a NLC are very well understood and the particle-wall interaction has been widely observed experimentally for a single particle immersed in a nematic cell [41, 45, 46], the properties of a single particle in a uniform NLC cell in the presence of an external electric field theoretically have not been fully addressed. Thus, it is of crucial importance to investigate the nature of a single particle in a uniform NLC cell in the presence of an external electric field.

## 8.2 Fréedericksz Transition in NLC

In a uniaxial nematic liquid crystal, the anisotropy of nematic phase is characterized by a symmetric and traceless tensor order parameter  $Q_{\alpha\beta}$  which can be written as

$$Q_{\alpha\beta} = S(n_\alpha n_\beta - \frac{1}{3}\delta_{\alpha\beta}). \quad (8.1)$$

Here  $n_\alpha$  and  $n_\beta$  are components of the director  $\mathbf{n}$ , which is a unit vector with the property  $\mathbf{n} = -\mathbf{n}$ , describing the direction along which the molecules are aligned. And  $S$  is the scalar order parameter that describes the degree of nematic order. It also represents how well the molecules are aligned along  $\mathbf{n}$ . If  $S$  equal to 0, there is no alignment, which means that the system is in an isotropic phase; If  $S$  equal to 1, it corresponds to a perfect alignment. When the director field  $\mathbf{n}(\mathbf{r})$  changes drastically due to the distortion from undeformed ground state in nematic liquid crystal, it costs elastic energy for the deviation of the director, which can be classified into three types, namely splay, twist and bend, making the Frank-Oseen free energy density for elastic distortions reads as [47]

$$f_{el} = \frac{1}{2}K_{11}(\nabla \cdot \mathbf{n})^2 + \frac{1}{2}K_{22}[\mathbf{n} \cdot (\nabla \times \mathbf{n})]^2 + \frac{1}{2}K_{33}[\mathbf{n} \times (\nabla \times \mathbf{n})]^2, \quad (8.2)$$

where  $K_{11}$ ,  $K_{22}$  and  $K_{33}$  are Frank elastic constants corresponding to splay, twist and bend constants, respectively. The bulk free energy of the nematic liquid crystal sample can be obtained by integrating  $f_{el}$  over the sample volume

$$F_{el} = \int f_{el} dV. \quad (8.3)$$

LC molecules are sensitive to weak external stimuli, such as electric field, magnetic field and light, due to the anisotropic property of NLC. The facile response to weak external stimuli results in the easy distortion of director field when a magnetic or an electric field is applied. If an external magnetic field is applied to the NLC, the following extra term should be added to the free energy

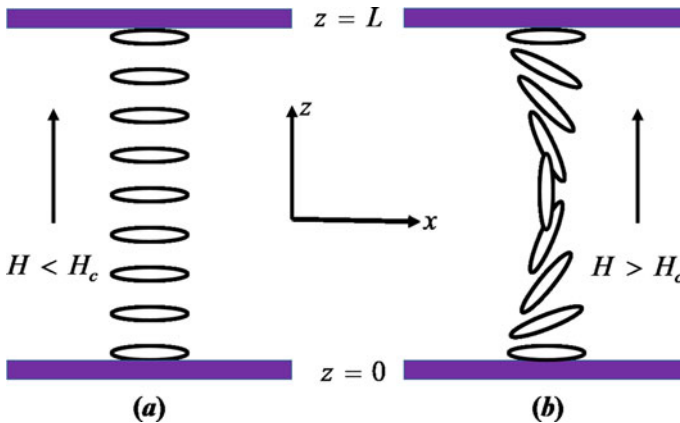
$$f_H = -\frac{1}{2}\Delta\chi(\mathbf{H} \cdot \mathbf{n})^2, \quad (8.4)$$

where  $\mathbf{H}$  is the magnetic field and  $\Delta\chi = \chi_{\parallel} - \chi_{\perp}$  is the diamagnetic anisotropy of the NLC, which can be positive or negative. Here  $\chi_{\parallel}$  and  $\chi_{\perp}$  are the two components of magnetic susceptibility for liquid crystal molecules when the magnetic field is applied. If  $\Delta\chi > 0$ , the molecules tend to align parallel to the direction of  $\mathbf{H}$ , while if  $\Delta\chi < 0$ , the molecules tend to align perpendicularly to the field direction. Analogously, if the external field is an electric one, then alternatively the additional free energy becomes

$$f_E = -\frac{1}{8\pi}\Delta\varepsilon(\mathbf{E} \cdot \mathbf{n})^2, \quad (8.5)$$

where  $\mathbf{E}$  is the electric field and  $\Delta\varepsilon = \varepsilon_{\parallel} - \varepsilon_{\perp}$  is the dielectric anisotropy of the NLC, which can be positive or negative as well. Here  $\varepsilon_{\parallel}$  and  $\varepsilon_{\perp}$  are the dielectric susceptibilities of the liquid crystal molecule parallel and perpendicular to the molecular long axis respectively. To illustrate how external fields alter the interactions of liquid crystal molecules, let us consider an NLC with thickness  $L$  sandwiched between two cell walls, and we choose the coordinate  $z$  axis normal to the cell walls where LC molecules are parallel to the  $x$  direction, as depicted in Fig. 8.1. Suppose a magnetic field  $\mathbf{H}$  is applied along the  $z$  direction. Then the director field deviating from the undeformed director  $\mathbf{n}_0 = (0,0,1)$  is given by

$$\mathbf{n}(z) = (\cos\theta(z), 0, \sin\theta(z)), \quad (8.6)$$



**Fig. 8.1** Sketch of Fréedericksz transition. The director  $\mathbf{n}$  is fixed in the  $x$  direction at the two plates of the cell, while the direction of the applied magnetic field  $H$  is perpendicular to the cell walls. **a** If  $H$  is below a certain critical field  $H_c$ , the NLC remains aligned in the  $x$  direction. **b** If  $H$  is above  $H_c$ , the NLC molecules start to try to realign along the  $z$  direction

where  $\theta$  is the angle between the director and  $z$  axis, satisfying the boundary condition  $\theta(0) = \theta(L) = 0$  due to the surface anchoring. The total free energy is then

$$F_{total} = \frac{1}{2} \int dz \left[ (K_{11} \cos^2 \theta(z) + K_{33} \sin^2 \theta(z)) \left( \frac{d\theta(z)}{dz} \right)^2 - \Delta\chi H^2 \sin^2 \theta(z) \right]. \quad (8.7)$$

To obtain the distribution  $\theta(z)$  corresponding to minimum total free energy, we calculate the functional derivative of the total free energy with respect to  $\theta(z)$ . Using one Frank constant approximation  $K_{11} = K_{33} = K$  leads to the Euler-Lagrange equation

$$\frac{d^2\theta(z)}{dz^2} + \frac{1}{\xi^2} \sin \theta(z) \cos \theta(z) = 0, \quad (8.8)$$

where  $\xi = \sqrt{K/(\Delta\chi H^2)}$  is called the magnetic coherence length. When the magnetic field is small (i.e.,  $L/\xi$  is small), the solution is  $\theta(z) = 0$ . However, the solutions vary as the magnetic field increases and exceeds a certain threshold. To address this problem, we assume the functional form of  $\theta(z)$  can be approximated by

$$\theta(z) = \theta_0 \sin\left(\frac{\pi z}{L}\right). \quad (8.9)$$

If  $\theta_0 \ll 1$ , substituting Eq. (8.9) into the total free energy of Eq. (8.7) and integrate over  $z = 0$  to  $L$ , we obtain

$$F_{total} = \frac{\pi^2 K}{4L} \theta_0^2 - \frac{\Delta\chi L H^2}{4} \theta_0^2 = \frac{\Delta\chi L}{4} (H_c^2 - H^2) \theta_0^2, \quad (8.10)$$

where  $H_c$  is the critical field

$$H_c = \frac{\pi}{L} \sqrt{\frac{K}{\Delta\chi}}. \quad (8.11)$$

Below the critical field (i.e.,  $H < H_c$ ), we have the solution  $\theta_z = 0$  throughout the cell, and the NLC remains aligned in the  $x$  direction (see Fig. 8.1a). If  $H > H_c$  the deviation of the director field takes place (see Fig. 8.1b) and such a order transition called the Fréedericksz transition.

When a microparticle is introduced into the NLC, the microparticle interacts with the surrounding liquid crystal primarily via surface anchoring. The resulting surface anchoring free energy can be expressed as an integral over the microparticle surface in the Rapini-Popular form [1, 48]

$$F_{anchoring} = \frac{1}{2} W \int (\mathbf{n} \cdot \mathbf{v})^2 d\mathbf{S} \quad (8.12)$$

where  $W$  is the anchoring coefficient and  $\mathbf{v}$  is a unit vector along the easy axis. Typically, the order of magnitude of  $W$  varies within the range  $10^{-4} \text{ mJ/m}^2 - 1 \text{ mJ/m}^2$ ,

where  $10^{-3} - 10^{-4}$  mJ/m<sup>2</sup> is considered as weak anchoring and  $1 - 10^{-1}$  mJ/m<sup>2</sup> is regarded as strong anchoring. In most cases the inclusion of particles into an NLC cell tends to create LC alignment singularities around the suspended substances, which in general are determined by surface anchoring conditions, particle size, boundary conditions, and external fields etc. [17, 49–52]. It has been widely accepted and confirmed that when a spherical particle is immersed in NLC, there are three possible types of defect configurations [53–55]. Dipole and quadrupolar configurations are usually seen around a spherical particle with strong vertical surface anchoring, whereas boojum defect is formed by a micro-sphere with tangential surface anchoring. In addition, recently B. Senyuk et al. assumed that conically degenerate boundary condition gives rise to the so-called elastic hexadecapole [56], and then Y. Zhou reported that the dipole-hexadecapole transformation can be achieved via tuning the preferred tilt angle of LC molecules anchoring on colloidal particle surface [57]. Through experimental observations it has been found that, when an external field is applied, there exists a transition between elastic dipole and quadrupolar configuration, which depends on particle size and surface anchoring strength [58–60].

As an application of the above theory, let us first of all consider a system that a particle is embedded in a uniform NLC without confinement. There are now two contributions to the free energy. The first contribution is the elastic deformation of the LC and can be accounted by the well known Frank-Oseen free energy. With one constant approximation  $K_{11} = K_{22} = K_{33} = K$ , the bulk deformation energy can be written as

$$F_b = \frac{K}{2} \int dV [(\nabla \cdot \mathbf{n})^2 + (\nabla \times \mathbf{n})^2]. \quad (8.13)$$

The second contribution is the surface anchoring free energy which is in the Rapini-Popula form Eq. (8.12), and the integration is over the particle surface. To determine the distribution of the director field  $\mathbf{n}(\mathbf{r})$  for a particle embedded in a NLC, the goal is solve the Euler-Lagrange equations arising from the variation of the total free energy  $F = F_b + F_{anchoring}$ . Unfortunately, the Euler-Lagrange equations with subjected boundary conditions at the surface of the particle and parallel boundary conditions at infinity are highly nonlinear, and analytical solutions are quite difficult to found. However, utilising the multipole expansion method similar to electrostatic [61], we can obtain analytic solutions for the director field far from the particle. We assume that the director field at infinite approach the undeformed director field  $\mathbf{n}_0 = (0,0,1)$  when there is no other confinement. The deviation of  $\mathbf{n}(\mathbf{r})$  from  $\mathbf{n}_0$  induced by the embedded particle is small at the large distance but not infinite, and  $\mathbf{n}(\mathbf{r}) \approx (\mathbf{n}_x, \mathbf{n}_y, 1)$ . Therefore, at large  $r$ , the nonlinear bulk free energy of deformation can be replaced by the harmonic free energy [53, 55]

$$F_{har} = \frac{K}{2} \int dV (\nabla n_\mu)^2 \quad (8.14)$$

with Euler-Lagrange equations of Laplace type

$$\nabla^2 n_\mu = 0. \quad (8.15)$$

Here  $n_\mu$  ( $\mu = x, y$ ) represents the components of the director field  $\mathbf{n}$  perpendicular to  $\mathbf{n}_0$ . Expanding the solutions into multipoles and we have the form of the solutions as follows [53, 55]

$$n_x = p \frac{x}{r^3} + 3c \frac{xz}{r^5}, \quad (8.16)$$

$$n_y = p \frac{y}{r^3} + 3c \frac{yz}{r^5}, \quad (8.17)$$

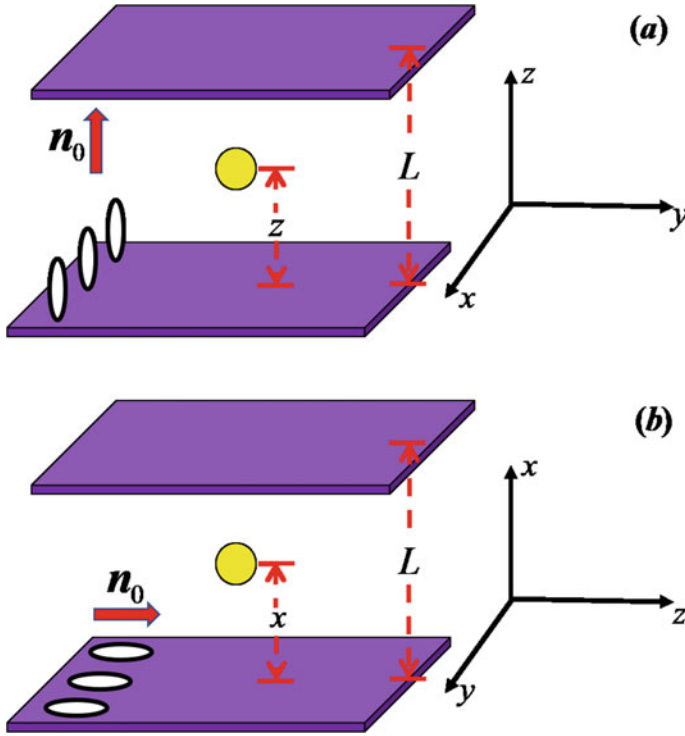
where  $p$  and  $c$  are the magnitude of the dipole and quadrupole moments respectively. If a particle is immersed in a NLC with confinement (i.e., NLC cell) and in the presence of external electric field, another surface anchoring free energy at the two plates of the cell and free energy arising from the applied electric field are need to be added. Thus our task is to minimize the more complicated total free energy functional. Unfortunately, it becomes more difficult to find the analytical solutions for the system of a particle suspended in a NLC cell in the presence of external field. Therefore, it is necessary to develop a phenomenological method to address this problem [40, 55], and this approach is introduced in the next section.

### 8.3 Theoretical Modeling

In order to introduce the phenomenological method mentioned above in details, we take the system that a spherical microparticle of radius  $r$  suspended in a NLC cell sandwiched between two parallel plates a distance  $L$  apart in the presence of an electric field as an example. The polarization of the particle is neglected compared with the influence of external field on the alignment of liquid crystal molecules. Figure 8.2 illustrates two systems schematically under external field with a homeotropic anchoring (Fig. 8.2a) and a homogeneous planar anchoring (Fig. 8.2b) respectively at the two cell walls. The suspended microparticle induces the director distortion and the director deviations  $n_\mu$  ( $\mu = x, y$ ) from the undeformed director field  $\mathbf{n}_0 = (0,0,1)$  are small at the region far from the microparticle. In order to use the same set of symbol subscripts ( $n_\mu$  ( $\mu = x, y$ )) in our theoretical modelling for the two surface anchoring conditions, two different coordinate frames are deliberately used here, as illustrated in Fig. 8.2a and b. Assuming  $\mathbf{n} \approx (n_x, n_y, 1)$  with one Frank constant approximation, the effective elastic energy for the system reads [42]

$$U_e = K \int d^3x \left[ \frac{(\nabla n_\mu)^2}{2} - \frac{k^2}{2} (\mathbf{e} \cdot \mathbf{n})^2 - 4\pi P(\mathbf{x}) \partial_\mu n_\mu - 4\pi C(\mathbf{x}) \partial_z \partial_\mu n_\mu \right], \quad (8.18)$$





**Fig. 8.2** Schematic representation of a microparticle of radius  $r$  suspended in a nematic liquid crystal cell with surface-to-surface distance  $L$  and **a** normal anchoring and **b** planar anchoring in the presence of an external electric field

where  $k^2 = (4\pi K)^{-1} \Delta \epsilon E^2$ ,  $P(\mathbf{x})$  and  $C(\mathbf{x})$  denote the dipole and the quadrupole moment densities respectively. When an electric field is applied along  $z$  axis, the Euler-Lagrange equation are given by [42]

$$\Delta n_\mu - k^2 n_\mu = 4\pi [\partial_\mu P(\mathbf{x}) - \partial_z \partial_\mu C(\mathbf{x})]. \tag{8.19}$$

When the external electric field is applied parallel to  $x$  axis, we have the Euler-Lagrange equations written as [42]

$$\Delta n_\mu + k^2 \delta_{x\mu} n_\mu = 4\pi [\partial_\mu P(\mathbf{x}) - \partial_z \partial_\mu C(\mathbf{x})]. \tag{8.20}$$

If the applied electric field is parallel to  $y$  axis, then the Euler-Lagrange equations are [42]

$$\Delta n_\mu + k^2 \delta_{y\mu} n_\mu = 4\pi [\partial_\mu P(\mathbf{x}) - \partial_z \partial_\mu C(\mathbf{x})]. \tag{8.21}$$

With Dirichlet boundary conditions  $n_\mu(\mathbf{s}) = 0$  on the two walls, the solution to Euler-Lagrange equations can be written as [42]

$$n_\mu(\mathbf{x}) = \int_V d^3 \mathbf{x}' G_\mu(\mathbf{x}, \mathbf{x}') [-\partial'_\mu P(\mathbf{x}') + \partial'_\mu \partial'_z C(\mathbf{x}')], \quad (8.22)$$

where  $G_\mu$  is the Green's function for  $n_\mu$ . Notice that here  $\mu$  in the integral does not follow Einstein summation notation.

## 8.4 Results and Discussions

### 8.4.1 Homeotropic Boundary Condition

#### 8.4.1.1 External Field Perpendicular to the Two Plates

Here we choose the coordinate  $z$  axis along the normal direction of the two cell walls where LC molecules are homeotropically anchored, as depicted in Fig. 8.2a). In the first case, when an electric is applied perpendicular to the two plates, i.e.,  $\mathbf{E} \parallel z$  in Fig. 8.2a, the corresponding Euler-Lagrange equations are written as Eq. (8.19). With Dirichlet boundary conditions  $n_\mu(z = 0) = n_\mu(z = L) = 0$ , the Green's function can be derived as [42]

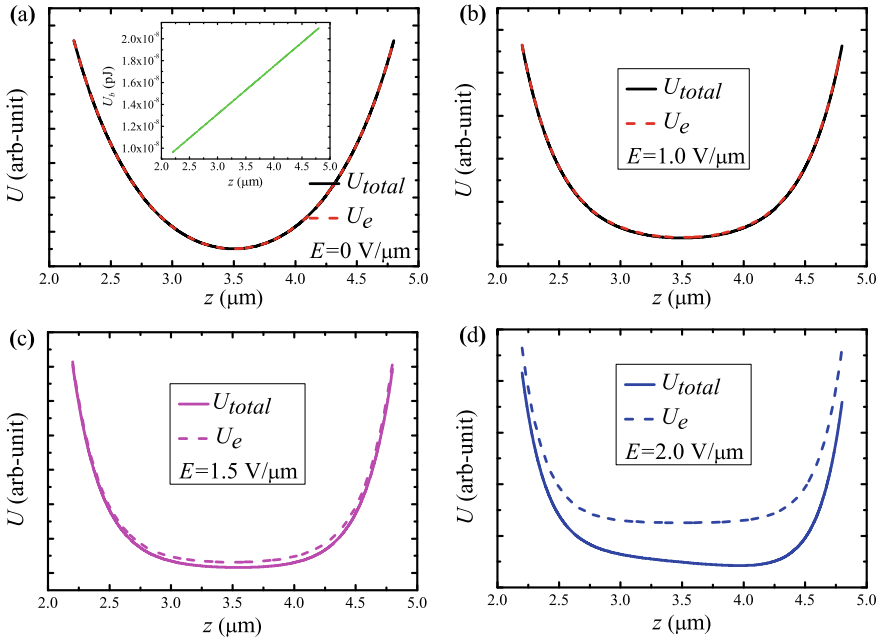
$$G_\mu(\mathbf{x}, \mathbf{x}') = \frac{4}{L} \sum_{n=1}^{\infty} \sum_{m=-\infty}^{\infty} e^{im(\varphi-\varphi')} \sin \frac{n\pi z}{L} \sin \frac{n\pi z'}{L} I_m(\lambda_n \rho_{<}) K_m(\lambda_n \rho_{>}). \quad (8.23)$$

Here  $\varphi$  and  $\varphi'$  are the azimuthal angles,  $z$  and  $z'$  are the positional coordinates,  $I_m$  and  $K_m$  are modified Bessel functions,  $\rho_{<}$  is the smaller one between  $\sqrt{x^2 + y^2}$  and  $\sqrt{x'^2 + y'^2}$ , and  $\lambda_n = [(n\pi/L)^2 + \Delta\epsilon E^2/4\pi K]^{1/2}$  with  $L$  the thickness of the NLC cell. Using the definition of self energy given in terms of Green's function[42]

$$U_{dd}^{self} = -2\pi K p^2 \partial_\mu \partial'_\mu H_\mu(\mathbf{x}, \mathbf{x}')|_{\mathbf{x}=\mathbf{x}'}, \quad (8.24)$$

where  $H_\mu(\mathbf{x}, \mathbf{x}') = G_\mu(\mathbf{x}, \mathbf{x}') - 1/|\mathbf{x} - \mathbf{x}'|$ , we can obtain the elastic energy  $U_e^I$  for an NLC cell with a microparticle suspended in the presence of an electric field. Besides the elastic energy, the gravitational potential  $U_g$  due to buoyant force should be considered as well, leading to a total energy written as

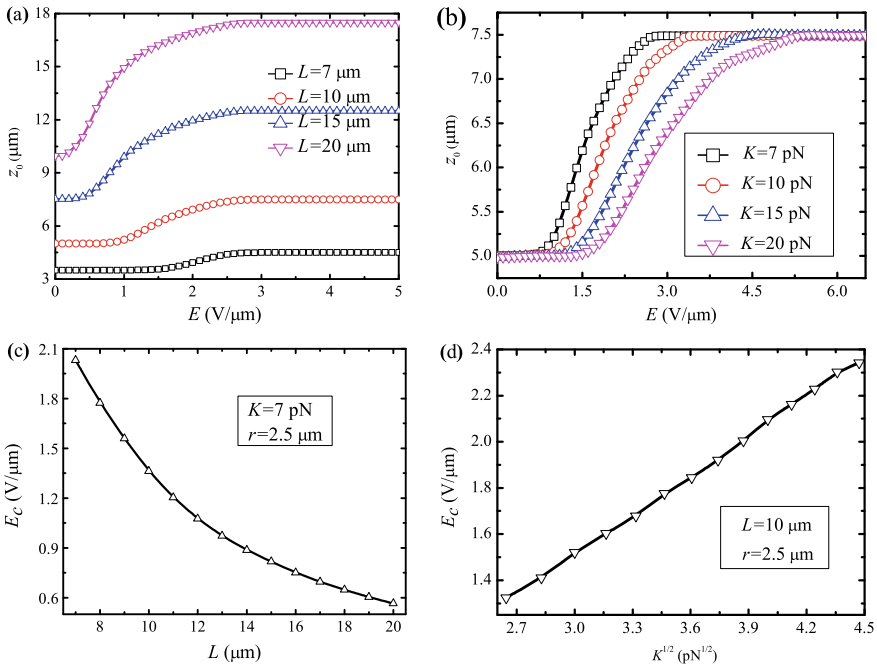
$$\begin{aligned} U_{total}^I &= U_e^I + U_g \\ &= -2\pi K p^2 \left[ -\frac{4}{L} \sum_{n=1}^{\infty} \lambda_n^2 \sin^2\left(\frac{n\pi z}{L}\right) K_0(\lambda_n \rho) + \frac{1}{\rho^3} \right]_{\rho \rightarrow 0} - \frac{4}{3} \pi r^3 (\rho_{LC} - \rho_{mp}) g z, \end{aligned} \quad (8.25)$$



**Fig. 8.3** Elastic energy and total energy as a function of microparticle position for different electric fields (0, 1.0, 1.5 and 2.0 V/ $\mu\text{m}$ ). Here we set the radius of microparticle and cell thickness as 2.2  $\mu\text{m}$  and 7  $\mu\text{m}$ , respectively

where  $r$  is the radius of microparticle,  $p = 2.04r^2$  is the magnitude of the equivalent dipole moment,  $\rho_{LC} - \rho_{mp}$  is the density difference between liquid crystal and microparticle,  $g = 9.8 \text{ m/s}^2$  is the gravitational acceleration, and  $z$  denotes the position of microparticle.

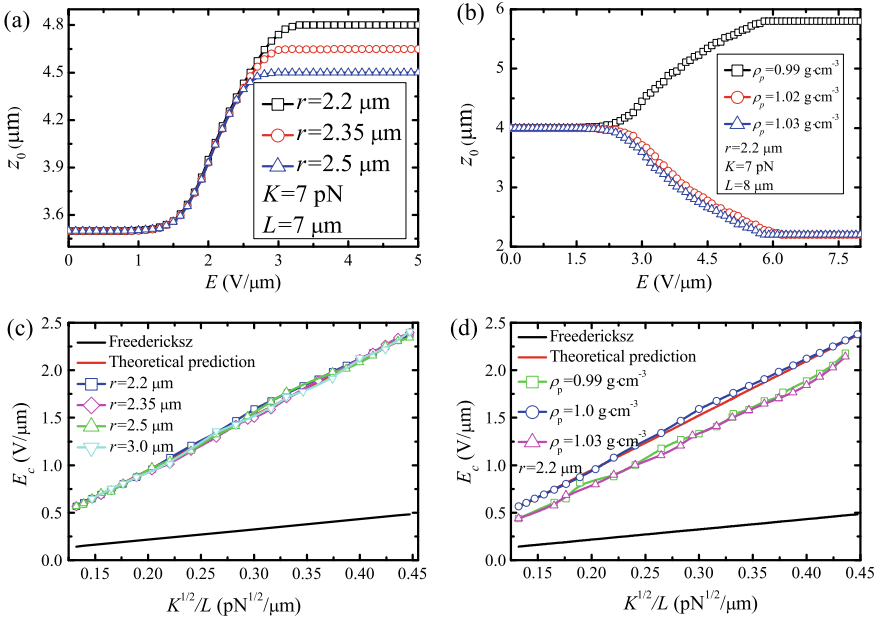
Based on the total energy obtained by Green’s function method, we now plot the profiles of total energy as a function of microparticle position for different electric field. Let us first consider the case  $\Delta\varepsilon > 0$ . The total energy and elastic energy against for four different electric field strengths are shown in Fig. 8.3. In the presence of a small external electric field, the total energy given by Eq. 8.25 overlaps the elastic energy  $U_e^I$  and remains symmetric, indicating that the interaction among LC molecules still dominates the system if the external field applied is not large enough to realign the LC molecules, especially in the region close to the midplane. Thus the contribution made by asymmetric gravitational potential is trivial and the microparticle in this case is still trapped within its midplane, as shown in Fig. 8.3b and c. However, as we increase the field applied, it tends to widen and flatten the bottom of the elastic potential well and that by contrast enlarges the relative contribution made by the asymmetric buoyant force to the total energy. As a result, the buoyant force will drive the microparticle with ease from midplane to a new equilibrium position (Fig. 8.3d). It is obvious that the sign of  $\rho_{LC} - \rho_{mp}$  determines the direction



**Fig. 8.4** Equilibrium position  $z_0$  in response to electric field for different **a** cell thicknesses (7, 10, 15 and 20  $\mu\text{m}$ ), where the Frank elastic constant and the radius of microparticle are set as  $K = 7$  pN and  $r = 2.5$   $\mu\text{m}$  respectively, and **b** Frank elastic constants (7, 10, 15 and 20 pN), where the cell thickness and the radius of microparticle are set as  $L = 10$   $\mu\text{m}$  and  $r = 2.5$   $\mu\text{m}$  respectively. These two figures show a positional transition occurring at an electric field threshold  $E_c$ , which depends on **c** cell thickness  $L$  and **d** Frank elastic constant  $\sqrt{K}$

of particle displacement with respect to the original equilibrium position. It seems that the interaction potential well around the midplane tends to be flattened due to the realignment of liquid crystal molecules made by the applied external field, which creates a “fast lane” in the vertical direction for the microparticle to move. It triggers a positional transition from the midplane, if driven by an asymmetric buoyant force, when such a fast lane is fast enough (weakens the elastic energy gradient).

Furthermore, in order to study the effect of the cell thickness and Frank constant on the critical electric value, we plot the equilibrium position against the applied electric field for different cell thicknesses (7, 10, 15 and 20  $\mu\text{m}$ ) and Frank elastic constants (7, 10, 15 and 20 pN), as shown in Fig. 8.4a and b. It is found that a positional transition occurs when the external field applied exceeds a threshold value. The thinner the cell thickness  $L$  is and the larger the Frank elastic constant  $K$  is, the larger the critical electric field is needed to trigger the transition, as shown in Fig. 8.4c and d. A deeper investigation shows that the critical value of electric field is inversely proportional to  $L$  and linearly proportional to  $\sqrt{K}$ , a Fréedericksz-like behavior.



**Fig. 8.5** Equilibrium position  $z_0$  for different **a** radii (2.2  $\mu\text{m}$ , 2.35  $\mu\text{m}$  and 2.5  $\mu\text{m}$ ) with  $K = 7$  pN and  $L = 7 \mu\text{m}$ ; **b** densities (0.99, 1.02 and 1.03  $\text{g} \cdot \text{cm}^{-3}$ ) of microparticle with  $K = 7$  pN and  $L = 8 \mu\text{m}$ , showing the same critical value  $E_c$  of electric field triggering positional transition. The dependence of  $E_c$  and  $\sqrt{K}/L$  for different **c** radii (2.2  $\mu\text{m}$ , 2.35  $\mu\text{m}$ , 2.5  $\mu\text{m}$  and 3.0  $\mu\text{m}$ ); **d** densities (0.99, 1.0 and 1.03  $\text{g} \cdot \text{cm}^{-3}$ ) of microparticle, obeying strictly a master curve given by theoretical prediction Eq. (8.26)

In order to gain more insight into the dynamic behaviors of the microparticle, we investigate the dependence of threshold value on various microparticle’s sizes and densities in Fig. 8.5. Figure 8.5a and b depict the equilibrium position against the applied electric field for different microparticle sizes and densities, where the overlapping of equilibrium position in Fig. 8.5a suggests that the critical electric value is almost independent of microparticle size. Whereas the symmetry of the equilibrium position of microparticle with density equal to 0.99  $\text{g} \cdot \text{cm}^{-3}$  and 1.03  $\text{g} \cdot \text{cm}^{-3}$  in Fig. 8.5b indicates that the slope of the master curve of critical electric value is nearly independent of the magnitude of microparticle density. Furthermore, to understand the dynamic behaviors of the microparticle, we plot the threshold value against  $\sqrt{K}/L$  to obtain a master curve, as shown in Fig. 8.5c and d, where a Fréedericksz curve (black) is also plotted. It is interesting to find that the critical electric field to trigger a positional transition for microparticle suspended in a NLC cell follows a Fréedericksz-like linear master curve, yet with a different slope. The existence of slightly difference instead of overlapping to each other for the equilibrium position of microparticle with density equal to 1.02  $\text{g} \cdot \text{cm}^{-3}$  and 1.03  $\text{g} \cdot \text{cm}^{-3}$  in Fig. 8.5b, leads to different intercepts of the Fréedericksz-like linear master curves for critical

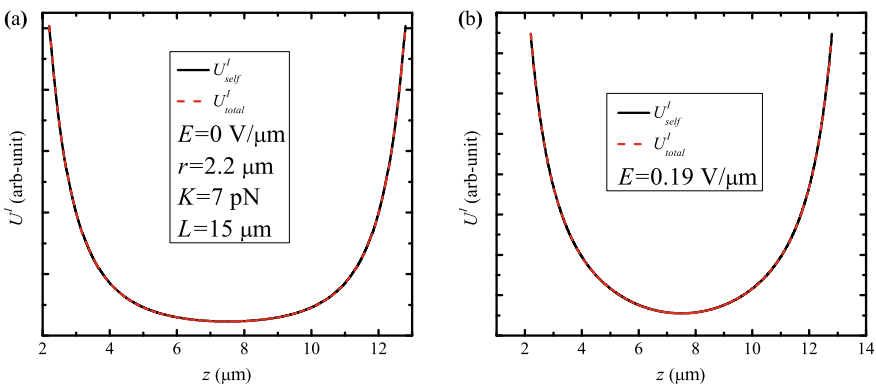
electric field in Fig. 8.5d. Obviously, the results indicate that the critical electric field for a positional transition to occur for a microparticle suspended in a NLC cell remains unchanged for different microparticle sizes and densities.

Moreover, by comparing the numerical calculation results with the Fréedericksz effect curve ( $\pi\sqrt{4\pi/|\Delta\varepsilon|}\sqrt{K}/L$ ) in Fig. 8.5c and d, it is surprising to find that the slope difference between them is by a factor of  $\sim 3\sqrt{\pi}$ . The additional energy contribution coming from the surface energy due to the introduction of microparticle is proportional to  $\pi$  (surface area). While on the other hand, the energy contribution made by external field is proportional to  $E^2$ , and that gives a critical value of external field proportional to  $\sqrt{\pi}$ , if the transition comes from the competition between equivalent surface energy due to the introduction of microparticle, and the Coulomb interaction due to the application of external field. More specifically, an explicit expression (where  $\mathcal{F}$  denotes the Fréedericksz effect)

$$E_c \simeq 3\sqrt{\pi}\mathcal{F} - \frac{1}{5} = 6\pi^2\sqrt{\frac{K}{|\Delta\varepsilon|L^2}} - \frac{1}{5} \tag{8.26}$$

for critical electric field can be proposed as a theoretical prediction. Such a prediction, as shown by straight line (red) in Fig. 8.5c and d, agrees very well for different radii (2.2, 2.35, 2.5, and 3.0  $\mu\text{m}$ ) and densities (0.99, 1.0 and 1.03  $\text{g} \cdot \text{cm}^{-3}$ ) of microparticle. This once again verifies the conclusion that the critical electric field is independent of microparticle size and density. The reason might lie in that in the present theoretical model, the microparticle is treated as a dipole in the far field expansion approximation.

In the case when  $\Delta\varepsilon < 0$ , the elastic energy and total energy as a function of microparticle position for two different electric field strengths is plotted as well, as shown in Fig. 8.6. Fig. 8.6a and b clearly show that the microparticle is trapped at



**Fig. 8.6** Elastic energy and total energy as a function of microparticle position for different electric fields **a** 0  $\text{V}/\mu\text{m}$  and **b** 0.19  $\text{V}/\mu\text{m}$ . Here the radius of microparticle, elastic constant and cell thickness are fixed at 2.2  $\mu\text{m}$ , 7 pN and 15  $\mu\text{m}$ , respectively

the midplane of the NLC cell, indicating that an application of external electric field does not trigger a positional phase transition. Moreover, in these two figures we also find that the electric field has a crucial impact on the shape of potential well of elastic energy and total energy. This is because when  $\mathbf{E}\parallel z$  and  $\Delta\varepsilon < 0$ , the realignment of liquid crystal molecules with the increase of the electric field narrows down the interaction potential well rather than flatten it (see Fig. 8.6), which creates a force directing toward the midplane much larger than the gravitational contribution and thus denies any positional transition.

### 8.4.1.2 External Field Parallel to the Two Plates

For the case of an electric field parallel to the two plates, i.e.,  $\mathbf{E}\parallel x$  in Fig. 8.2a, the Euler-Lagrange equations for  $n_x$  and  $n_y$  are written as Eq. (8.20). With Dirichlet boundary conditions  $n_\mu(z = 0) = n_\mu(z = L) = 0$ , the related Green's functions  $G_x$  and  $G_y$  are given by

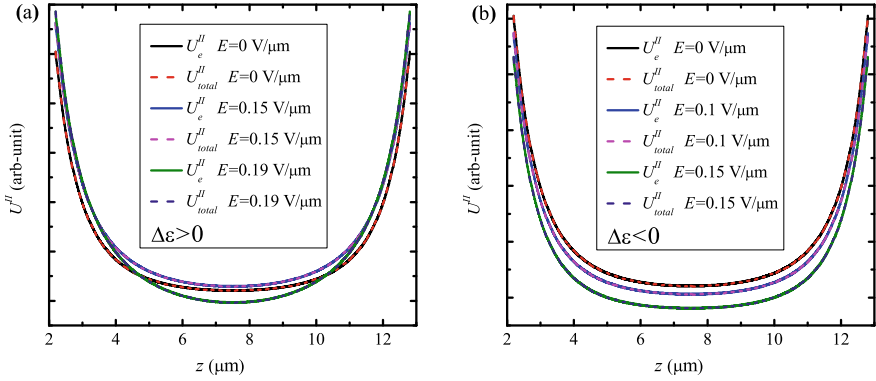
$$\begin{aligned}
 G_x(\mathbf{x}, \mathbf{x}') &= \frac{4}{L} \sum_{n=1}^{\infty} \sum_{m=-\infty}^{\infty} e^{im(\varphi-\varphi')} \sin \frac{n\pi z}{L} \sin \frac{n\pi z'}{L} I_m(v_n \rho_{<}) K_m(v_n \rho_{>}), \\
 G_y(\mathbf{x}, \mathbf{x}') &= \frac{4}{L} \sum_{n=1}^{\infty} \sum_{m=-\infty}^{\infty} e^{im(\varphi-\varphi')} \sin \frac{n\pi z}{L} \sin \frac{n\pi z'}{L} I_m(\mu_n \rho_{<}) K_m(\mu_n \rho_{>}), \quad (8.27)
 \end{aligned}$$

where  $v_n = [(n\pi/L)^2 - \Delta\varepsilon E^2/4\pi K]^{1/2}$  and  $\mu_n = n\pi/L$ . In analogue to the previous case, we can obtain the elastic energy  $U_e^{II}$  and thereby the total energy  $U_{total}^{II}$  is written as

$$\begin{aligned}
 U_{total}^{II} &= U_e^{II} + U_g \\
 &= -2\pi K p^2 \left[ -\frac{2}{L} \sum_{n=1}^{\infty} \sin^2\left(\frac{n\pi z}{L}\right) (\alpha_n + \beta_n) + \frac{1}{\rho^3} \right]_{\rho \rightarrow 0} - \frac{4}{3} \pi r^3 (\rho_{LC} - \rho_{mp}) g z, \quad (8.28)
 \end{aligned}$$

where  $\alpha_n = v_n^2 K_0(v_n \rho) + \mu_n^2 K_0(\mu_n \rho)$ , and  $\beta_n = v_n^2 K_2(v_n \rho) - \mu_n^2 K_2(\mu_n \rho)$ .

In order to seek for the effect of electric field on the equilibrium position of microparticle, plots for the elastic energy and total energy against the microparticle position for different electric fields are presented in Fig. 8.7. When  $\Delta\varepsilon > 0$ , the narrows down of the potential well of elastic energy and total energy with the increase of electric field in Fig. 8.7a indicates that the microparticle tends to be trapped in the midplane of the NLC cell. When  $\Delta\varepsilon < 0$ , the increases of electric field tends to deepen and narrows down (which can hardly tell) the potential well of elastic energy and total energy, as illustrated in Fig. 8.7b. Due to the deep potential well, the microparticle is trapped in the midplane of the NLC cell regardless of the sign of the molecular dielectric anisotropy, indicating that an application of external elec-



**Fig. 8.7** Elastic energy and total energy as a function of the microparticle position for different electric fields with **a**  $\Delta\epsilon > 0$  and **b**  $\Delta\epsilon < 0$ . Here the radius of microparticle, elastic constant and cell thickness are fixed at  $2.2 \mu\text{m}$ ,  $7 \text{ pN}$  and  $15 \mu\text{m}$ , respectively

tric field, however large it is, can not trigger a positional transition. This can be understood by considering the fact that the molecular long (short) axes tend to align along the direction of applied electric field as  $\Delta\epsilon > 0$  ( $\Delta\epsilon < 0$ ). As we increase the field applied, the interaction potential is found to be narrowed down and deepened, corresponding to a strong midplane-directing restoring force. Therefore for the homeotropic boundary condition, the positional transition occurs only in an NLC cell with positive molecular dielectric anisotropy when the external electric field is applied along the undeformed director field.

## 8.4.2 Planar Boundary Condition

### 8.4.2.1 External Field Perpendicular to the Two Plates

Now we turn to the situation that LC molecules are horizontally anchored on the two cell walls and an electric field is applied vertically to the two plates, i.e.,  $\mathbf{E} \parallel x$ , as depicted in Fig. 8.2b. The Euler-Lagrange equations are given by Eq. (8.20), and the corresponding Green's functions  $G_x$  and  $G_y$  read as [42]

$$G_x(\mathbf{x}, \mathbf{x}') = \frac{4}{L} \sum_{n=1}^{\infty} \sum_{m=-\infty}^{\infty} e^{im(\varphi-\varphi')} \sin \frac{n\pi x}{L} \sin \frac{n\pi x'}{L} I_m(v_n \rho_{<}) K_m(v_n \rho_{>}),$$

$$G_y(\mathbf{x}, \mathbf{x}') = \frac{4}{L} \sum_{n=1}^{\infty} \sum_{m=-\infty}^{\infty} e^{im(\varphi-\varphi')} \sin \frac{n\pi x}{L} \sin \frac{n\pi x'}{L} I_m(\mu_n \rho_{<}) K_m(\mu_n \rho_{>}), \quad (8.29)$$

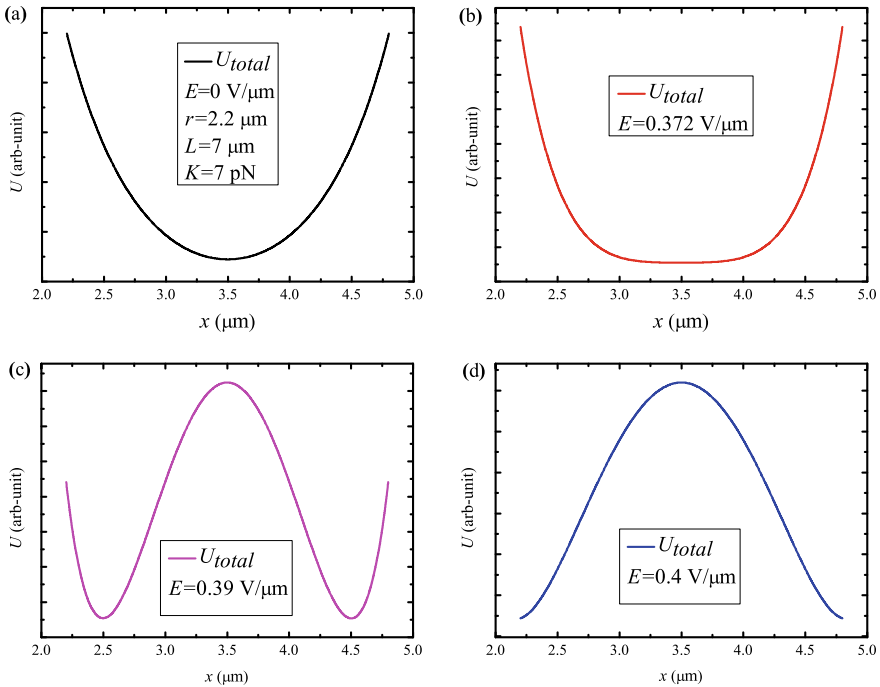
with  $v_n$  and  $\mu_n$  identical to those in Eq. (8.27). Similarly, the elastic energy  $U_e^{III}$  can be obtained and the total energy  $U_{total}^{III}$  can be derived as



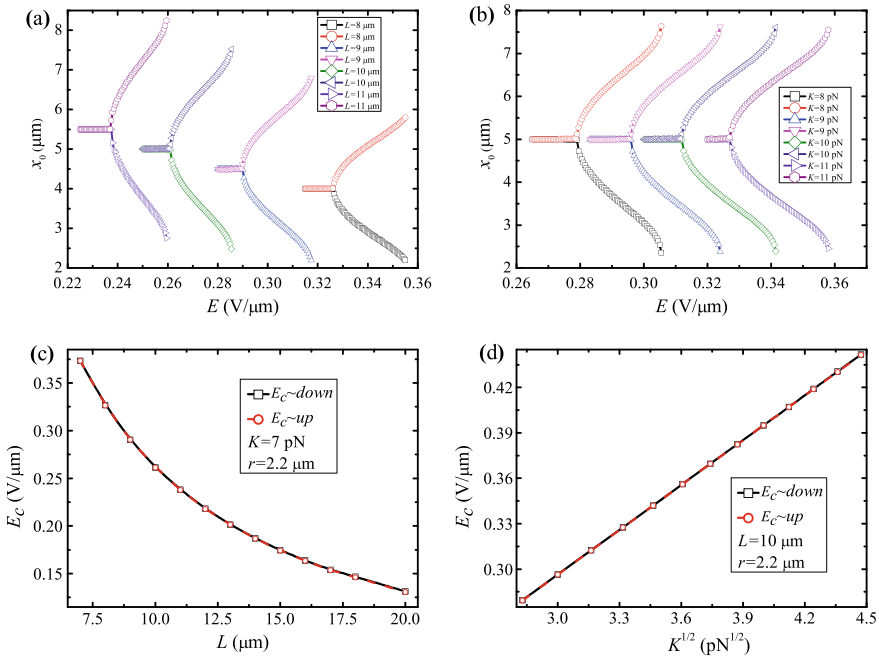
$$\begin{aligned}
 U_{total}^{III} &= U_e^{III} + U_g \\
 &= -2\pi K p^2 \left[ \frac{4}{L} \sum_{n=1}^{\infty} \mu_n^2 \left[ \cos^2\left(\frac{n\pi x}{L}\right) K_0(v_n \rho) - \frac{1}{2} \sin^2\left(\frac{n\pi x}{L}\right) (K_0(\mu_n \rho) - K_2(\mu_n \rho)) \right] + \frac{1}{\rho^3} \right]_{\rho \rightarrow 0} \\
 &\quad - \frac{4}{3} \pi r^3 (\rho_{LC} - \rho_{mp}) g x,
 \end{aligned} \tag{8.30}$$

where  $x$  denotes the vertical position of the microparticle.

In this case, let us first consider a positive dielectric anisotropy  $\Delta\varepsilon > 0$ . Intriguingly, a significant feature is observed regarding the profile of total energy as a function of microparticle position for four different electric fields, as illustrated in Fig. 8.8. In the presence of small field (below the critical electric value), Fig. 8.8a and b show that the interaction potential well around the midplane tends to be flattened in this region due to the realignment of liquid crystal molecules made by the increment of external electric field. However, when the electric field rises beyond the threshold value, there exists two symmetric equilibrium positions for the suspended microparticle (see Fig. 8.8c and d). Which one the microparticle shifts to is decided by the perturbation stemming from the asymmetric buoyant force, i.e. by the density difference between NLC and microparticle ( $\rho_{LC} - \rho_{mp}$ ). Notably, the total energy now is almost equal to the elastic energy due to the fact that the gravitational



**Fig. 8.8** Total energy profile as a function of the suspended microparticle position for an NLC cell with planar anchoring in the presence of different electric fields perpendicular to the two plates

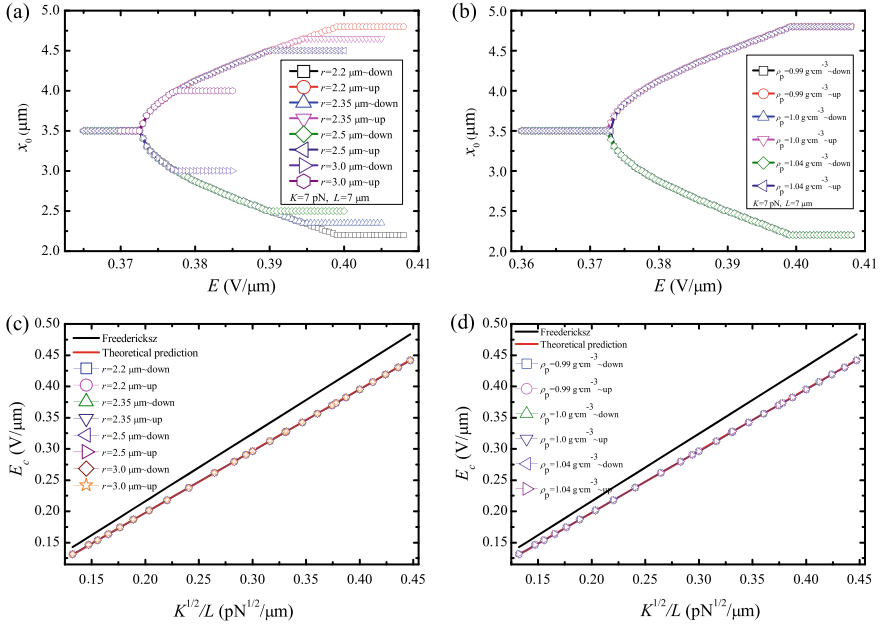


**Fig. 8.9** Equilibrium position  $x_0$  in response to electric field for different **a** cell thicknesses (8, 9, 10 and 11  $\mu\text{m}$ ) with Frank elastic constant  $K = 7$  pN and the radius of microparticle  $r = 2.2$   $\mu\text{m}$ , and **b** Frank elastic constants (8, 9, 10 and 11 pN) with cell thickness  $L = 10$   $\mu\text{m}$  and radius of microparticle  $r = 2.2$   $\mu\text{m}$ . It is shown that the electric field threshold  $E_c$  at which a positional transition occurs, depends on **c** cell thickness  $L$  and **d** Frank elastic constant  $\sqrt{K}$

contribution is much smaller in contrast to the elastic one, generating the depths of the two local minimums in Fig. 8.8c (and Fig. 8.8d) nearly equal to each other.

To probe the influence of cell thickness and Frank constant on the critical field value, we plot the equilibrium position of the suspended microparticle against the applied electric field for different cell thicknesses (8, 9, 10 and 11  $\mu\text{m}$ ) and Frank elastic constants (8, 9, 10 and 11 pN), as shown in Fig. 8.9a and b, where a positional transition occurs at some electric field threshold values and there exist two bistable equilibrium positions when the external field applied exceeds the critical value. The thinner the cell thickness  $L$  is and the larger the Frank elastic constant  $K$  is, the larger the critical electric field is needed to trigger the positional transition, as shown in Fig. 8.9c and d. A more deeper investigation shows that the critical value of electric field is inversely proportional to  $L$  and linearly proportional to  $\sqrt{K}$ , a Fréedericksz-like behavior.

As a following step, we examine whether the critical electric value is correlated with the size and density of the microparticle. Surprisingly, Fig. 8.10a and b show that the plots of the equilibrium position of suspended microparticle against the applied electric field for different microparticle sizes and densities overlap each other,



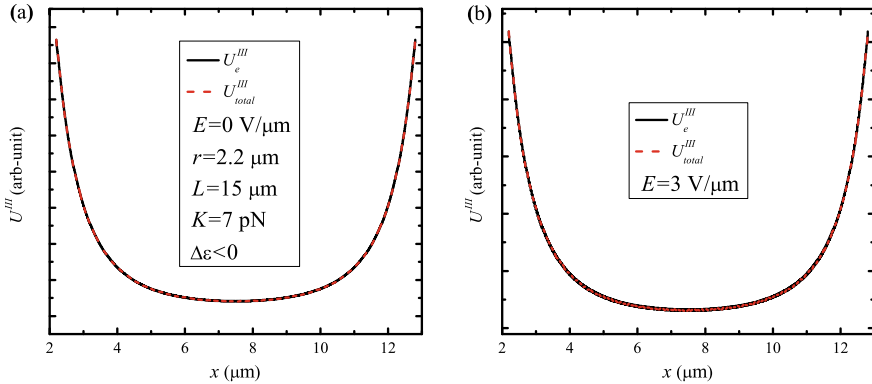
**Fig. 8.10** Equilibrium position  $x_0$  for different **a** radii (2.2  $\mu\text{m}$ , 2.35  $\mu\text{m}$ , 2.5  $\mu\text{m}$  and 3.0  $\mu\text{m}$ ); **b** densities (0.99, 1.0 and 1.04  $\text{g} \cdot \text{cm}^{-3}$ ) of a microparticle with  $K = 7 \text{ pN}$  and  $L = 7 \mu\text{m}$ , showing the same critical value  $E_c$  of electric field triggering positional transition. The dependence of  $E_c$  on  $\sqrt{K}/L$  for different **c** radii (2.2  $\mu\text{m}$ , 2.35  $\mu\text{m}$ , 2.5  $\mu\text{m}$  and 3.0  $\mu\text{m}$ ); **d** densities (0.99, 1.0 and 1.04  $\text{g} \cdot \text{cm}^{-3}$ ) of the microparticle, obeying strictly a master curve which can be given by the theoretical prediction Eq. (8.31)

suggesting that the critical electric value is independent of or negligibly depends on microparticle size and density. To gain more insight into the dynamic behaviors of the microparticle, we further plot the threshold value against  $\sqrt{K}/L$  in Fig. 8.10c and d, where a Fréedericksz curve (black) is shown as well. It is interesting to find that the critical electric field to trigger a positional transition for a microparticle suspended in an NLC cell follows a Fréedericksz-like linear master curve with slightly different slopes, a universal one also valid for different microparticle sizes and densities.

By comparing the numerical calculation results with the Fréedericksz transition ( $\pi \sqrt{4\pi} / |\Delta\varepsilon| \sqrt{K}/L$ ) in Fig. 8.10c and d, we found that the slope difference between them is by a prefactor of  $\sim 0.915$ , and that enables us to propose a theoretical prediction for the critical electric field

$$E_c \simeq 0.915\mathcal{F}, \tag{8.31}$$

where  $\mathcal{F}$  denotes the Fréedericksz effect. Such a prediction, as shown by straight line (red) in Fig. 8.10c and d, agrees very well for different radii (2.2, 2.35, 2.5,



**Fig. 8.11** Elastic energy and total energy profile as a function of the suspended microparticle position for an NLC cell with planar anchoring in the presence of different electric fields perpendicular to the two plates. Here the radius of microparticle, elastic constant and cell thickness are set as 2.2  $\mu\text{m}$ , 7 pN and 15  $\mu\text{m}$ , respectively

and 3.0  $\mu\text{m}$ ) and densities (0.99, 1.0 and 1.04  $\text{g} \cdot \text{cm}^{-3}$ ) of microparticle. Due to the mathematical difficulty, we still don't know how to derive 0.915 analytically.

In the case when  $\Delta\epsilon < 0$ , to determine whether the positional transition phenomenon takes place or not, we plot the elastic energy and total energy against microparticle position for different electric field, as depicted in Fig. 8.11. The comparison of Fig. 8.11a and b shows that the potential well of elastic energy and total energy are narrowed down with the increase of electric field, indicating that the suspended microparticle is trapped at the midplane of the NLC cell, which can be predicted by the profile change of the total energy potential well due to application of an external electric field in the vertical direction ( $x$  direction in Fig. 8.2b). The short axes of liquid crystal molecules tend to align along the electric field, a result leading to the narrowing of total potential well and thereby generating strong restoring force acting on the suspended microparticle. Therefore, in the case of a microparticle suspended in an NLC cell with planar anchoring condition in the presence of an external electric field applied perpendicular to the two plates, the positional transition triggered by the electric field occurs only under the condition of positive molecular dielectric anisotropy.

#### 8.4.2.2 External Field Parallel to the Two Plates but Perpendicular to the Anchoring Direction

Now let us consider the case when the electric field applied is parallel to the two plates but perpendicular to the anchoring direction, i.e.,  $\mathbf{E} \parallel y$  in Fig. 8.2b, the Euler-Lagrange equations can be given by Eq. (8.21), with their corresponding Green's functions  $G_x$  and  $G_y$  written as

$$\begin{aligned}
G_x(\mathbf{x}, \mathbf{x}') &= \frac{4}{L} \sum_{n=1}^{\infty} \sum_{m=-\infty}^{\infty} e^{im(\varphi-\varphi')} \sin \frac{n\pi x}{L} \sin \frac{n\pi x'}{L} I_m(\mu_n \rho_{<}) K_m(\mu_n \rho_{>}), \\
G_y(\mathbf{x}, \mathbf{x}') &= \frac{4}{L} \sum_{n=1}^{\infty} \sum_{m=-\infty}^{\infty} e^{im(\varphi-\varphi')} \sin \frac{n\pi x}{L} \sin \frac{n\pi x'}{L} I_m(\nu_n \rho_{<}) K_m(\nu_n \rho_{>}), \quad (8.32)
\end{aligned}$$

with the same  $\nu_n$  and  $\mu_n$  as those in Eq. (8.29). In a similar way, the total energy  $U_{total}^{IV}$  is given by

$$\begin{aligned}
U_{total}^{IV} &= U_e^{IV} + U_g \\
&= -2\pi K \rho^2 \left[ \frac{4}{L} \sum_{n=1}^{\infty} \mu_n^2 \cos^2\left(\frac{n\pi x}{L}\right) K_0(\mu_n \rho) - \frac{2}{L} \sum_{n=1}^{\infty} \nu_n^2 \sin^2\left(\frac{n\pi x}{L}\right) (K_0(\nu_n \rho) - K_2(\nu_n \rho)) + \frac{1}{\rho^3} \right]_{\rho \rightarrow 0} \\
&\quad - \frac{4}{3} \pi r^3 (\rho_{LC} - \rho_{mp}) g x, \quad (8.33)
\end{aligned}$$

where  $U_e^{IV}$  is the elastic energy.

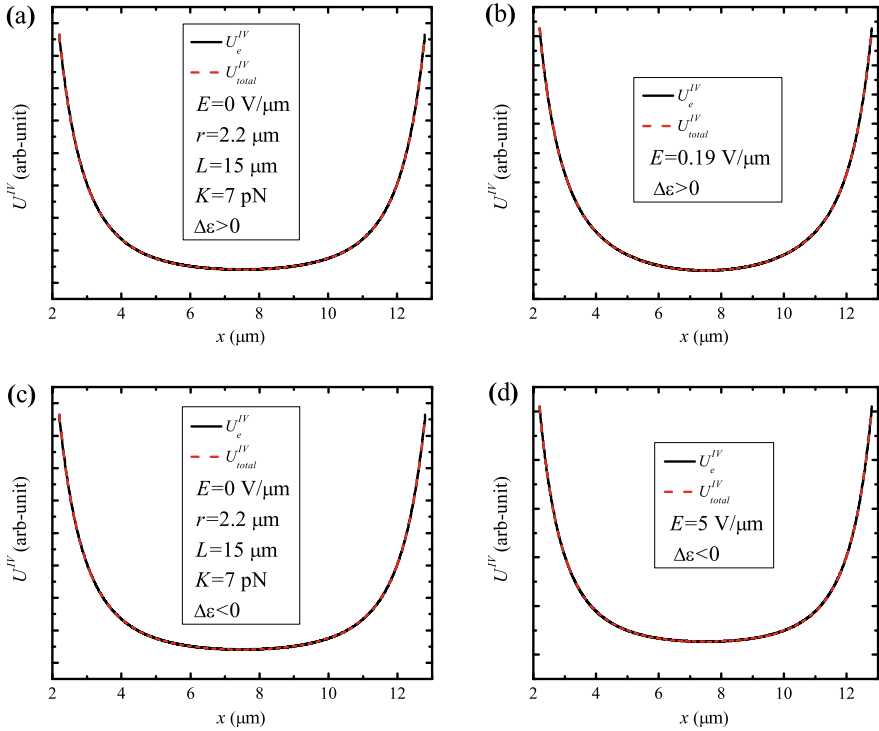
Similar to the previous cases, to probe the influence of electric field on the equilibrium position of microparticle, plots for the elastic energy and total energy against the microparticle position for different electric field are presented in Fig. 8.12. When  $\Delta\varepsilon > 0$ , the potential well of elastic energy and total energy in Fig. 8.12b are narrowed down as compared with that in Fig. 8.12a. When  $\Delta\varepsilon < 0$ , Fig. 8.12c and d show the same trend. Those results suggest that no matter  $\Delta\varepsilon > 0$  or  $\Delta\varepsilon < 0$ , the microparticle is always trapped at the midplane of the NLC cell regardless of the magnitude of the electric field applied, indicating that no positional transition occurs. The reason lies in that the realignment of the liquid crystal molecules in the presence of the external electric field does not flatten the interaction potential well substantially enough so as to decrease its corresponding equivalent restoring force on the microparticle to a small magnitude, with which the asymmetric gravitational force becomes competitive.

#### 8.4.2.3 External Field Parallel to the Two Plates and the Anchoring Direction

Finally, we consider an NLC cell in the presence of an electric field parallel to the two plates and the anchoring direction as well, i.e.,  $\mathbf{E} \parallel z$  in Fig. 8.2b. Given the corresponding Euler-Lagrange equations Eq. (8.19), the Green's functions are [42]

$$G_\mu(\mathbf{x}, \mathbf{x}') = \frac{4}{L} \sum_{n=1}^{\infty} \sum_{m=-\infty}^{\infty} e^{im(\varphi-\varphi')} \sin \frac{n\pi x}{L} \sin \frac{n\pi x'}{L} I_m(\lambda_n \rho_{<}) K_m(\lambda_n \rho_{>}), \quad (8.34)$$

with the same  $\lambda_n$  as that in Eq. (8.23). Similarly, the total energy  $U_{total}^V$  can be derived as

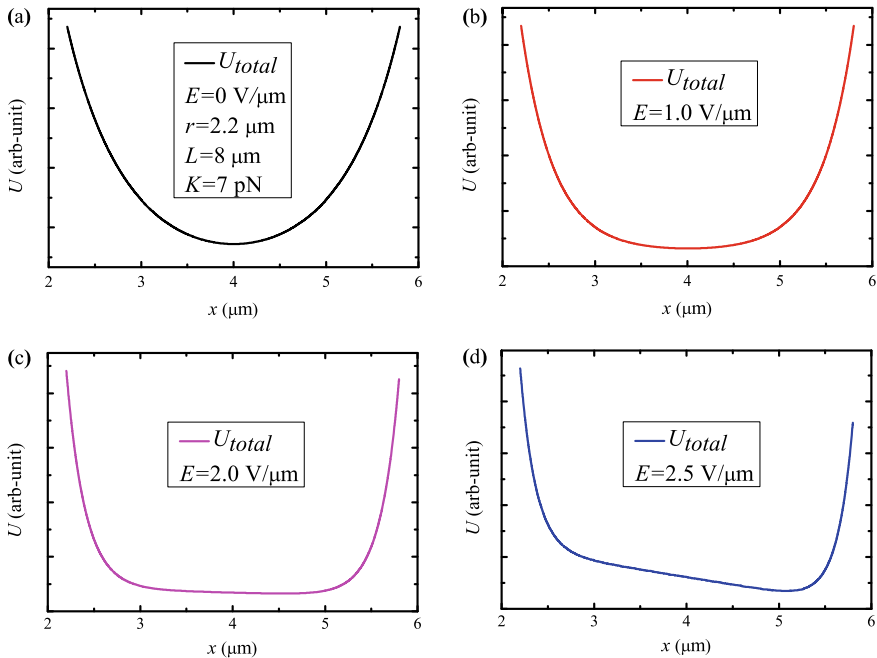


**Fig. 8.12** Elastic energy and total energy as a function of microparticle position for **a**  $\Delta\varepsilon > 0$  and  $E = 0 \text{ V}/\mu\text{m}$ ; **b**  $\Delta\varepsilon > 0$  and  $E = 0.19 \text{ V}/\mu\text{m}$ ; **c**  $\Delta\varepsilon < 0$  and  $E = 0 \text{ V}/\mu\text{m}$ ; **d**  $\Delta\varepsilon < 0$  and  $E = 5 \text{ V}/\mu\text{m}$ . Here the radius of microparticle, elastic constant and cell thickness are fixed at  $2.2 \mu\text{m}$ ,  $7 \text{ pN}$  and  $15 \mu\text{m}$ , respectively

$$\begin{aligned}
 U_{total}^V &= U_e^V + U_g \\
 &= -2\pi Kp^2 \left[ \frac{4}{L} \sum_{n=1}^{\infty} \left(\frac{n\pi}{L}\right)^2 \cos^2\left(\frac{n\pi x}{L}\right) K_0(\lambda_n \rho) - \frac{2}{L} \sum_{n=1}^{\infty} \lambda_n^2 \sin^2\left(\frac{n\pi x}{L}\right) (K_0(\lambda_n \rho) - K_2(\lambda_n \rho)) + \frac{1}{\rho^3} \right]_{\rho \rightarrow 0} \\
 &\quad - \frac{4}{3} \pi r^3 (\rho_{LC} - \rho_{mp}) g x, \tag{8.35}
 \end{aligned}$$

with  $U_e^V$  the elastic energy.

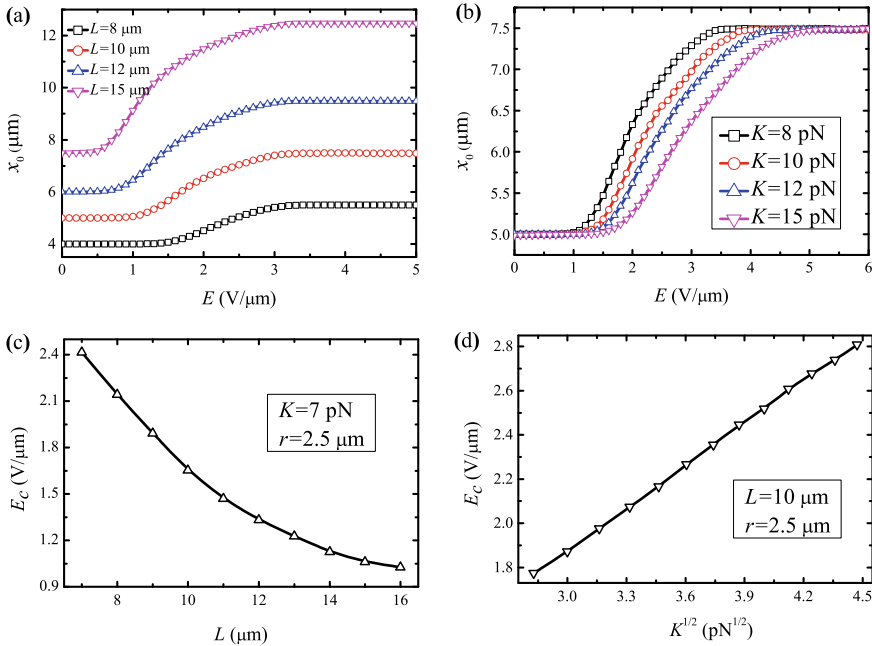
Given a positive molecular dielectric anisotropy, namely  $\Delta\varepsilon > 0$ , we can plot, as shown in Fig. 8.13, the total energy profile as a function of the suspended microparticle position for four chosen electric fields. In the presence of a small external field, the total energy profile remains symmetric, indicating that the elastic interaction among LC molecules dominates the LC alignment, especially in the region close to the midplane. Thus the contribution made by asymmetric gravitational potential is trivial if compared with elasticity and the suspended microparticle will be trapped within its midplane, as demonstrated in Fig. 8.13a and b. While as the electric field is increased, it is found that it tends to widen and flatten the bottom of the elastic poten-



**Fig. 8.13** Total energy profile as a function of the suspended microparticle position for an NLC cell with planar anchoring in the presence of four chosen electric fields parallel to the two plates and the anchoring direction as well

tial well, which equivalently by contrast amplifies the relative contribution made by the asymmetric buoyant force to the total energy of the NLC cell. As a result, the buoyant force will drive the microparticle with ease from the midplane to a new equilibrium position (see Fig. 8.13c and d). It is apparent that the sign of  $\rho_{LC} - \rho_{mp}$  determines the direction of the microparticle displacement. It looks very much like that the bottom of the interaction potential well around the midplane is “pressed due to the realignment of liquid crystal molecules made by the applied external field, which creates a “fast lane” along the vertical direction in the cell for the suspended microparticle to migrate. Once such a “fast lane” constructed by the external field in the cell reaches a critical value of “smoothness” (corresponding to a weakened elastic energy gradient), driven by the asymmetric buoyant force, it triggers a positional transition for the suspended microparticle from the midplane to its new equilibrium position.

In order to study the influence of cell thickness and Frank constant on the critical value of electric field, plots for the equilibrium position for the suspended microparticle against the applied electric field for different cell thicknesses (8, 10, 12 and 15  $\mu\text{m}$ ) and Frank elastic constants (8, 10, 12 and 15 pN) are presented in Fig. 8.14a and b, where it is found that a positional transition occurs when the external field applied exceeds a threshold value. It is also shown that the thinner the cell thickness

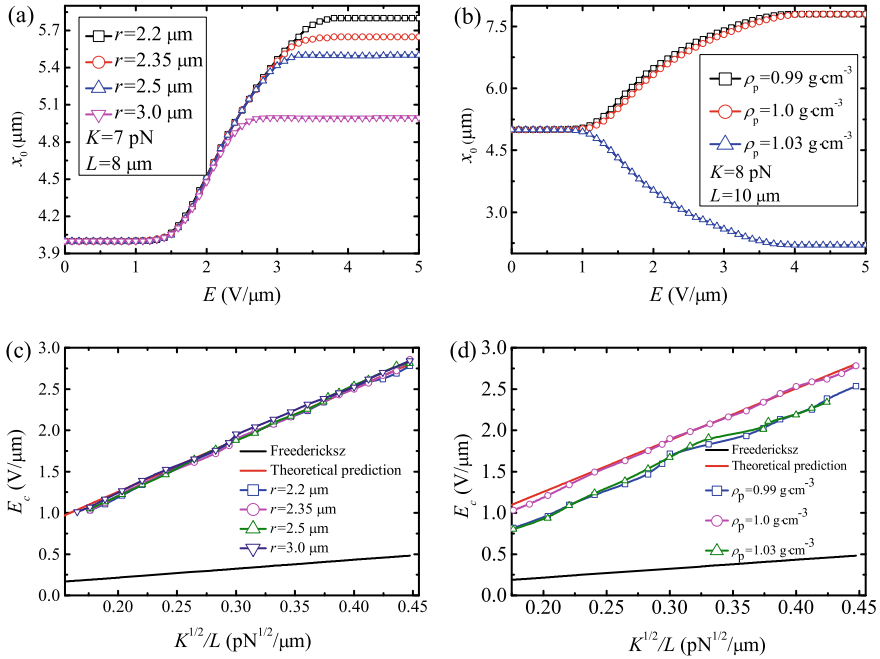


**Fig. 8.14** Equilibrium position  $x_0$  in response to electric field for different **a** cell thicknesses (8, 10, 12 and 15  $\mu\text{m}$ ), where the Frank elastic constant and the radius of microparticle are set as  $K = 7 \text{ pN}$  and  $r = 2.5 \mu\text{m}$ , and **b** Frank elastic constants (8, 10, 12 and 15  $\text{pN}$ ), where the cell thickness and the radius of microparticle are set as  $L = 10 \mu\text{m}$  and  $r = 2.5 \mu\text{m}$ . The electric field threshold  $E_c$  depends on **c** cell thickness  $L$  and **d** Frank elastic constant  $\sqrt{K}$

$L$  is and the larger the Frank elastic constant  $K$  is, the larger the critical electric field is needed to trigger the positional transition. The further study of electric field threshold (see Fig. 8.14c and d) shows that it seems to be inversely proportional to cell thickness  $L$  and proportional to the root square of Frank elastic constant  $K$ , a behavior similar to the field threshold of Fréedericksz phase transition.

In a similar way to the previous sections, the dependence of the threshold value on microparticles size and density is also investigated. Figure 8.15a and b depict the equilibrium position against the applied electric field for different microparticle sizes and densities, where the overlapping of equilibrium position in Fig. 8.15a suggests that the critical electric value is almost independent of microparticle size. Whereas the symmetry of the equilibrium position of microparticle with density equal to  $0.99 \text{ g} \cdot \text{cm}^{-3}$  and  $1.03 \text{ g} \cdot \text{cm}^{-3}$  in Fig. 8.15b indicates that the slope of the master curve of critical electric value is nearly independent of the magnitude of equivalent microparticle density. To gain more insight into the dynamic behaviors of the microparticle, the threshold value is plotted against  $\sqrt{K}/L$  in Fig. 8.15c and d, where a Fréedericksz transition curve (black) is shown as well. The existence of slightly difference instead of overlapping to each other for the equilibrium position





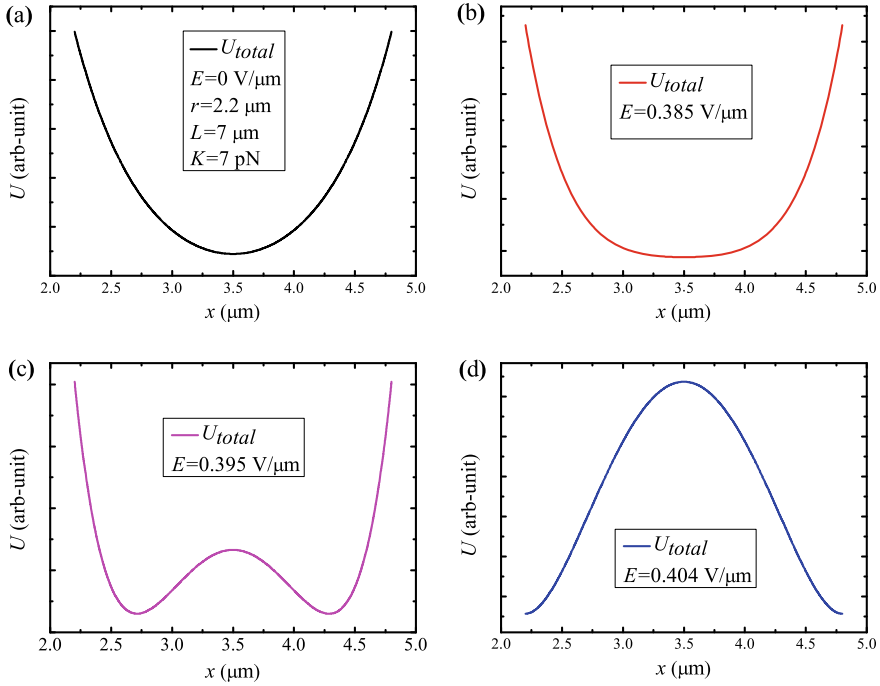
**Fig. 8.15** **a** Equilibrium position  $x_0$  for different radii of microparticle with  $K = 7$  pN and  $L = 8$   $\mu\text{m}$ , showing the same critical value  $E_c$  of electric field triggering positional transition. **b**  $K = 8$  pN and  $L = 10$   $\mu\text{m}$ . The dependence of  $E_c$  and  $\sqrt{K}/L$  for different **c** radii (2.2  $\mu\text{m}$ , 2.35  $\mu\text{m}$ , 2.5  $\mu\text{m}$  and 3.0  $\mu\text{m}$ ); **d** densities (0.99, 1.0 and 1.03  $\text{g} \cdot \text{cm}^{-3}$ ) of microparticle, obeying strictly a master curve given by theoretical prediction Eq. (8.36)

of microparticle with density equal to 0.99  $\text{g} \cdot \text{cm}^{-3}$  and 1.0  $\text{g} \cdot \text{cm}^{-3}$  in Fig. 8.15b, leads to different intercepts of the Fréedericksz-like linear master curves for critical electric field in Fig. 8.15d. Like before, the critical electric field for a positional transition to occur for a microparticle suspended in a NLC cell remains unchanged for different microparticle sizes and densities.

Similarly, a contrast between the numerical calculation results and the traditional Fréedericksz transition curve ( $\pi \sqrt{4\pi/|\Delta\varepsilon|} \sqrt{K}/L$ ) in Fig. 8.15c and d shows that the slope difference between them is by a prefactor of  $\sim 5.8$ . More specifically, an explicit expression

$$E_c \simeq 5.8\mathcal{F} - 0.08 = 5.8\pi \sqrt{\frac{4\pi K}{|\Delta\varepsilon|L^2}} - 0.08 \tag{8.36}$$

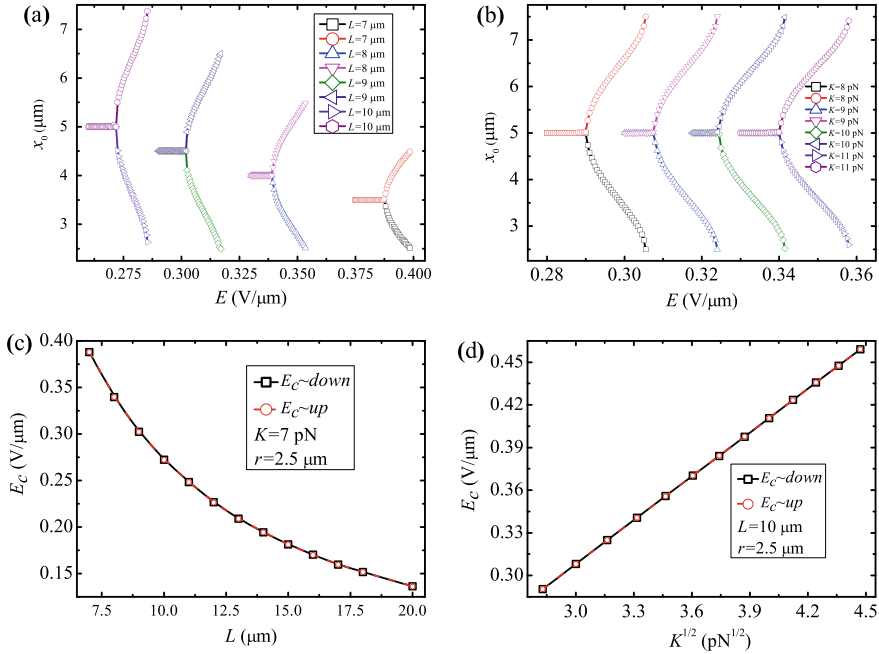
for critical electric field can be proposed as a theoretical prediction. Such a prediction, as shown by straight line (red) in Fig. 8.15c and d, agrees very well for different radii (2.2, 2.35, 2.5, and 3.0  $\mu\text{m}$ ) and densities (0.99, 1.0 and 1.03  $\text{g} \cdot \text{cm}^{-3}$ ) of microparticle. This once again verifies the conclusion that the critical electric field is



**Fig. 8.16** Total energy profile as a function of the suspended microparticle position for different external electric fields

independent of microparticle size, of which the reason might lie in that in the present theoretical model, the microparticle is approximately treated as a dipole in the far field expansion.

As for the case  $\Delta\varepsilon < 0$  when the external field applied parallel to both the two plates and the anchoring direction, i.e.,  $\mathbf{E} \parallel \mathbf{z}$  in Fig. 8.2b, a bistable equilibrium state structure is found as the electric field exceeds a threshold value, as illustrated in Fig. 8.16. In the small-field region, the external field applied tends to, first of all, flatten the bottom of potential well, as shown in Fig. 8.16a and b. Further increase of external field will change the one-state potential structure to a bistable one. As the gravitational contribution to the total energy is still negligibly small compared to the elastic one, one sees no involvement of gravitational force to the determination of the critical value of positional transition for the microparticle in the NLC cell. Thus, the positional transition in this case does not come from the competition between the gravitational force and the equivalent elastic force but rather purely from the bistable local minimum of the elastic potential, as shown in Fig. 8.16c and d. Nevertheless the asymmetric gravitational force still plays a very important role in determining the direction of microparticle motion (up or down) by acting as a small but significant

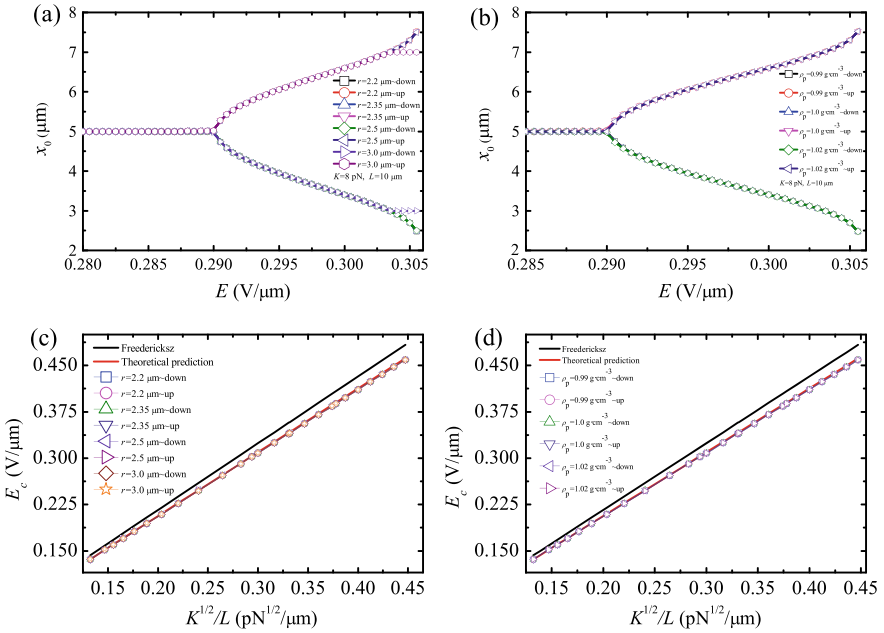


**Fig. 8.17** Equilibrium position  $x_0$  in response to electric field for different **a** cell thicknesses (7, 8, 9 and 10  $\mu\text{m}$ ), where the Frank elastic constant and the radius of microparticle are set as  $K = 7$  pN and  $r = 2.5$   $\mu\text{m}$ , and **b** Frank elastic constants (8, 9, 10 and 11 pN), where the cell thickness and the radius of microparticle are set as  $L = 10$   $\mu\text{m}$  and  $r = 2.5$   $\mu\text{m}$ . It is shown that a positional transition occurs at electric field threshold  $E_c$ , which depends on **c** cell thickness  $L$  and **d** Frank elastic constant  $\sqrt{K}$

perturbation, or more precisely, by the sign of buoyant force (the sign of  $\rho_{LC} - \rho_{mp}$ ). Therefore, the magnitude of the asymmetric gravitational force in this case is trivial but not its sign.

In order to understand how cell thickness and Frank elastic constant affect the critical value of electric field, we plot equilibrium position against the applied electric field for different cell thicknesses (7, 8, 9 and 10  $\mu\text{m}$ ) and Frank elastic constants (8, 9, 10 and 11 pN), as shown in Fig. 8.17a and b, where a bifurcation of equilibrium position is found due to the bistable state structure of elastic potential and a positional transition occurs when the external field applied reaches a threshold value. Additionally, a Fréedericksz-like behavior is shown in Fig. 8.17c and d. As observed, the thinner the cell thickness  $L$  is and the larger the Frank elastic constant  $K$  is, the larger the critical electric field is needed to trigger the positional transition, which corresponding to that the critical value of electric field is inversely proportional to  $L$  and linearly proportional to  $\sqrt{K}$ .

Finally, in order to gain more insights into the physics hidden behind the dynamic behaviors of microparticle, it is worthwhile to evaluate whether the critical electric



**Fig. 8.18** Equilibrium position  $x_0$  for different **a** radii (2.2  $\mu\text{m}$ , 2.35  $\mu\text{m}$ , 2.5  $\mu\text{m}$  and 3.0  $\mu\text{m}$ ); **b** densities (0.99, 1.0 and 1.02  $\text{g} \cdot \text{cm}^{-3}$ ) of microparticle with  $K = 8 \text{ pN}$  and  $L = 10 \mu\text{m}$ , showing the same critical value  $E_c$  of electric field triggering positional transition. The dependence of  $E_c$  and  $\sqrt{K}/L$  for different **c** radii (2.2  $\mu\text{m}$ , 2.35  $\mu\text{m}$  and 3.0  $\mu\text{m}$ ); **d** densities (0.99, 1.0 and 1.02  $\text{g} \cdot \text{cm}^{-3}$ ) of microparticle, obeying strictly a master curve given by theoretical prediction Eq. (8.37)

value is correlated with the size and density of the microparticle. The dependence of the equilibrium position on the applied electric field for different microparticle sizes and densities is shown in Fig. 8.18a and b, where the strict overlapping of equilibrium position in the figures implies that the critical electric value is, as shown in the previous section, independent of microparticle size and density. For a better understanding of the dynamic behaviors of the microparticle, we further plot the threshold value against  $\sqrt{K}/L$  in Fig. 8.18c and d, with a Fréedericksz transition curve (black) shown as well. It is found that the critical electric field triggering a positional transition for a microparticle suspended in a NLC cell follows a Fréedericksz master curve irrelevant to microparticle size and density.

More precisely, by comparing the numerical calculation results with the Fréedericksz effect curve ( $\pi \sqrt{4\pi/|\Delta\varepsilon|} \sqrt{K}/L$ ) in Fig. 8.18c and d, it is found that the slope difference between them is by a prefactor of  $\sim 3/\pi$ , leading to a proposed theoretical prediction for the critical electric field. Such a prediction, as shown by straight line (red) in Fig. 11c and d, agrees very well for different radii (2.2, 2.35, 2.5, and 3.0  $\mu\text{m}$ ) and densities (0.99, 1.0 and 1.02  $\text{g} \cdot \text{cm}^{-3}$ ) of microparticle.

**Table 8.1** Formation of a vertical fast “lane” for positional transition to occur (+) and not to occur (–) for a microparticle suspended in an NLC cell in the presence of an external electric field

Anchoring	Molecular dielectric anisotropy	Field direction		
		$\mathbf{E} \perp$ Plates	$\mathbf{E} \parallel$ Plates	
Homeotropic	$\Delta\varepsilon > 0$	+	–	
	$\Delta\varepsilon < 0$	–	–	
Planar	$\Delta\varepsilon > 0$	+/bistable	$\mathbf{E} \perp$ Anchoring	$\mathbf{E} \parallel$ Anchoring
	$\Delta\varepsilon < 0$		–	+
			–	+ /bistable

$$E_c \simeq \frac{3}{\pi} \mathcal{F} = 6 \sqrt{\frac{\pi K}{|\Delta\varepsilon| L^2}} \tag{8.37}$$

Based on the discussions in the sections above, it is quite obvious that the external electric field applied enhances the existing anisotropy of distortion generated by the boundaries of the NLC cell shaped by the movable suspended microparticle and the two parallel walls. It looks like there exists an anisotropic movable “bubble” surrounding the suspended microparticle, created by the external field and the boundary conditions combined. Inside the “bubble” along the vertical direction a fast “lane” will be constructed once the external field applied reaches a critical value. The electric field threshold is a signal to complete the construction and a “key” to switching on the use of the fast “lane”, wobbling the “bubble” along the vertical direction, and thereby tune the motion of the microparticle inside, which has been proved to be a positional transition [44]. Interestingly, this kind of motion can be found in the some SiFi novels picturing one of the possible tactics for intergalactic travel in the future by moving a planetary object via wobbling the space-time around it, which is supported by general relativity. After a thorough discussion of all the conditions combined to create such a wobbling “bubble” in a NLC cell in the presence of an external electric field, we come up with a table for a positional transition to occur in such a system, as shown in Table 8.1. It is found in the Table that out of the ten combinations of field direction, molecular dielectric anisotropy, and anchoring feature, only four shows the possible occurrence of positional transition. Moreover, for a nematic liquid crystal cell with planar surface alignment, a bistable equilibrium structure for the transition is found when the direction of applied electric field is (a) perpendicular to the two plates of the cell with positive molecular dielectric anisotropy, or (b) parallel to both the two plates and the anchoring direction of the cell with negative molecular dielectric anisotropy.

## 8.5 Conclusion

In summary, using the Green's function method, the total energy for a microparticle suspended in an NLC cell in the presence of an external electric field is calculated. It is found that with the application of the external electric field, it is possible to create an anisotropic bubble around the microparticle with a vertical fast "lane" for the microparticle to move from the midplane to a new equilibrium position. Such a new equilibrium position is decided via a competition between the buoyant force and the effective force built upon the microparticle inside the "lane". The threshold value of external field, which triggers positional transition under appropriate conditions of surface anchoring feature, field direction and molecular dielectric anisotropy, depends on thickness  $L$  and Frank elastic constant  $K$  and slightly on the microparticle size and density, in a Fréedericksz-like manner as coined by the authors before, but by a factor. For an NLC cell with planar surface alignment, a bistable equilibrium structure for the transition is found when the direction of the applied electric field is (a) perpendicular to the cell wall with positive molecular dielectric anisotropy, and (b) parallel to the undeformed director field  $\mathbf{n}_0$  of the NLC cell with negative molecular dielectric anisotropy. When the electric field applied is parallel to the two plates and perpendicular to the anchoring direction, the microparticle suspended in NLC will be trapped in the midplane, regardless of the sign of the molecular dielectric anisotropy. Explicit formulae proposed for the critical electric field agrees extremely well with the numerical calculation.

## References

1. de Gennes, P.G., Prost, J.: The Physics of Liquid Crystals, 2nd edn. Clarendon Press, Oxford, UK (1993)
2. Jákli, A., Lavrentovich, O.D., Selinger, J.V.: Physics of liquid crystals of bent-shaped molecules. *Rev. Mod. Phys.* **90**, 045004 (2018)
3. Daniel, J.C., Audebert, R.: Small Volumes and Large Surfaces: The World of Colloids in Soft Matter Physics edited by M. Williams (Springer-Verlag, Berlin Heidelberg, Daoud and C.E (1999)
4. Chaikin, P.M., Lubensky, T.C.: Principles of Condensed Matter Physics, Cambridge University Press, CambridgeCambridgeCambridge, UK (2000)
5. Comiskey, B., Albert, J., Yoshizawa, H., Jacobson, J.: *Nature* **394**, 253 (1998)
6. Wang, Z., Zhe, J.: *Chip* **11**, 1280 (2011)
7. Araki, T., Buscaglia, M., Bellini, T., Tanaka, H.: *Nat. Mater.* **10**, 303 (2011)
8. Smalyukh, I.I.: *Annu. Rev. Condens. Matter Phys.* **9**, 207 (2018)
9. Kim, Y.-K., Wang, X., Mondkar, P., Bukusoglu, E., Abbott, N.: *Nature* **557**, 539 (2018)
10. Nance, E.A., Woodworth, G.F., Sailor, K.A., Shih, T.-Y., Xu, Q., Swaminathan, G., Xiang, D., Eberhart, C., Hanes, J.: *Sci. Transl. Med.* **4**, 149ra119 (2012)
11. Woltman, S.J., Jay, G.D., Crawford, G.P.: *Nat. Mater.* **6**, 929 (2007)
12. Poulin, P., Cabuil, V., Weitz, D.A.: *Phys. Rev. Lett.* **79**, 4862 (1997)
13. Vilfan, M., Osterman, N., Čopič, M., Ravnik, M., Žumer, S., Kotar, J., Babič, D., Poberaj, I.: *Phys. Rev. Lett.* **101**, 237801 (2008)
14. Ognysta, U., Nych, A., Nazarenko, V., Muševič, I., Škarabot, M., Ravnik, M., Žumer, S., Poberaj, I., Babič, D.: *Phys. Rev. Lett.* **100**, 217803 (2008)

15. Škarabot, M., Ravnik, M., Žumer, S., Tkalec, U., Poberaj, I., Babič, D., Osterman, N., Muševič, I.: *Phys. Rev. E* **77**, 031705 (2008)
16. Ryzhkova, A.V., Škarabot, M., Muševič, I.: *Phys. Rev. E* **91**, 042505 (2015)
17. Lapointe, C.P., Mason, T.G., Smalyukh, I.I.: *Science* **326**, 1083 (2009)
18. Lapointe, C.P., Hopkins, S., Mason, T.G., Smalyukh, I.I.: *Phys. Rev. Lett.* **105**, 178301 (2010)
19. Ognysta, U.M., Nych, A.B., Uzunova, V.A., Pergamenschik, V.M., Nazarenko, V.G., Škarabot, M., Muševič, I.: *Phys. Rev. E* **83**, 041709 (2011)
20. Kim, S.-J., Lee, B.-K., Kim, J.-H.: *Liq. Cryst.* **43**, 1589 (2016)
21. Andrienko, D., Tasinkevych, M., Patricio, P., Allen, M.P., Teloda Gama, M.M.: *Phys. Rev. E* **68**, 051702 (2003)
22. Izaki, K., Kimura, Y.: *Phys. Rev. E* **87**, 062507 (2013)
23. Araki, T., Nagura, J.: *Phys. Rev. E* **95**, 012706 (2017)
24. Conklin, C., Tovkach, O.M., Viñals, J., Calderer, M.C., Golovaty, D., Lavrentovich, O.D., Walkington, N.J.: *Phys. Rev. E* **98**, 022703 (2018)
25. Muševič, I.: *Nematic liquid-crystal colloids. Materials* **11**, 24 (2018)
26. Calderon, F.L., Stora, T., Mondain Monval, O., Poulin, P., Bibette, J.: *Phys. Rev. Lett.* **72**, 2959 (1994)
27. Yada, M., Yamamoto, J., Yokoyama, H.: *Phys. Rev. Lett.* **92**, 185501 (2004)
28. Takahashi, K., Ichikawa, M., Kimura, Y.: *Phys. Rev. E* **77**, 020703(R) (2008)
29. Škarabot, M., Ryzhkova, A.V., Muševič, I.: *J. Mol. Liq.* **267**, 384 (2018)
30. Pishnyak, O.P., Tang, S., Kelly, J.R., Shiyonovskii, S.V., Lavrentovich, O.D.: *Phys. Rev. Lett.* **99**, 127802 (2007)
31. Pishnyak, O.P., Shiyonovskii, S.V., Lavrentovich, O.D.: *J. Mol. Liq.* **164**, 132 (2011)
32. Pagès, Josep M., Ignés-Mullol, Jordi, Sagués, Francesc: *Phys. Rev. Lett.* **122**, 198001 (2019)
33. Atzin, N., Guzmán, O., Gutiérrez, O.: *Phys. Rev. E* **97**, 062704 (2018)
34. Denniston, C., Orlandini, E., Yeomans, J.M.: *Phys. Rev. E* **63**, 056702 (2001)
35. Changizrezaei, S., Denniston, C.: *Phys. Rev. E* **99**, 052701 (2019)
36. Ravnik, M., Žumer, S.: *Liq. Cryst.* **36**, 1201 (2009)
37. Ravnik, M.: *Liq. Cryst. Today* **20**, 77 (2011)
38. Tasinkevych, M., Silvestre, N.M., da Gama, M.M.T.: *New J. Phys.* **14**, 073030 (2012)
39. Seyednejad, S.R., Mozaffari, M.R., Ejtehadi, M.R.: *Phys. Rev. E* **88**, 012508 (2013)
40. Chernyshuk, S.B., Lev, B.I.: *Phys. Rev. E* **81**, 041701 (2010)
41. Chernyshuk, S.B., Lev, B.I.: *Phys. Rev. E* **84**, 011707 (2011)
42. Chernyshuk, S.B., Tovkach, O.M., Lev, B.I.: *Phys. Rev. E* **85**, 011706 (2012)
43. D'Adamo, G., Marenduzzo, D., Micheletti, C., Orlandini, E.: *Phys. Rev. Lett.* **114**, 177801 (2015)
44. Xiao, K., Chen, X., Wu, C.X.: *Phys. Rev. Res.* **1**, 033041 (2019)
45. Kim, S.-J., Kim, J.-H.: *Soft Matter* **10**, 2664 (2014)
46. Lee, B.-K., Kim, S.-J., Lev, B., Kim, J.-H.: *Phys. Rev. E* **95**, 012709 (2017)
47. Frank, F.C.: *On the theory of liquid crystals. Faraday Discuss* **25**, 19–28 (1958)
48. Rapini, A., Papoular, M.: *J. Phys. (Paris), Colloq.* **30**, C4–54 (1969)
49. Stark, H.: *Phys. Rep.* **351**, 387 (2001)
50. Stark, H.: *Phys. Rev. E* **66**, 032701 (2002)
51. Wang, Y., Zhang, P., Chen, J.Z.Y.: *Phys. Rev. E* **96**, 042702 (2017)
52. Yao, X., Zhang, H., Chen, J.Z.Y.: *Phys. Rev. E* **97**, 052707 (2018)
53. Poulin, P., Stark, H., Lubensky, T.C., Weitz, D.A.: *Science* **275**, 1770 (1997)
54. Poulin, P., Weitz, D.A.: *Phys. Rev. E* **57**, 626 (1998)
55. Lubensky, T.C., Petey, D., Currier, N., Stark, H.: *Phys. Rev. E* **57**, 610 (1998)
56. Senyuk, B., Puls, O., Tovkach, O.M., Chernyshuk, S.B., Smalyukh, I.I.: *Nat. Commun.* **7**, 10659 (2016)
57. Zhou, Y., Senyuk, B., Zhang, R., Smalyukh, I.I., de Pablo, J.J.: *Nat. Commun.* **10**, 1000 (2019)
58. Ruhwandl, R.W., Terentjev, E.M.: *Phys. Rev. E* **54**, 5204 (1996)
59. Ruhwandl, R.W., Terentjev, E.M.: *Phys. Rev. E* **56**, 5561 (1997)
60. Loudet, J.C., Poulin, P.: *Phys. Rev. Lett.* **87**, 165503 (2001)
61. Jackson, J.D.: *Classical Electrodynamics*, 3rd edn. Wiley, New York (1999)