Springer Complexity

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

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

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

The three major book publication platforms of the Springer Complexity program are the monograph series "Understanding Complex Systems" focusing on the various applications of complexity, the "Springer Series in Synergetics", which is devoted to the quantitative theoretical and methodological foundations, and the "SpringerBriefs in Complexity" which are concise and topical working reports, case-studies, surveys, essays and lecture notes of relevance to the field. In addition to the books in these two core series, the program also incorporates individual titles ranging from textbooks to major reference works.

Springer Series in Synergetics

Founding Editor: H. Haken

The Springer Series in Synergetics was founded by Herman Haken in 1977. Since then, the series has evolved into a substantial reference library for the quantitative, theoretical and methodological foundations of the science of complex systems.

Through many enduring classic texts, such as Haken's *Synergetics and Information and Self-Organization*, Gardiner's *Handbook of Stochastic Methods*, Risken's *The Fokker Planck-Equation* or Haake's *Quantum Signatures of Chaos*, the series has made, and continues to make, important contributions to shaping the foundations of the field.

The series publishes monographs and graduate-level textbooks of broad and general interest, with a pronounced emphasis on the physico-mathematical approach.

More information about this series at http://www.springer.com/series/712

Peter Schuster

Stochasticity in Processes

Fundamentals and Applications to Chemistry and Biology

Springer

Peter Schuster
Institut für Theoretische Chemie
Universität Wien
Wien, Austria

ISSN 0172-7389 ISSN 2198-333X (electronic)
Springer Series in Synergetics
ISBN 978-3-319-39500-5 ISBN 978-3-319-39502-9 (eBook)
DOI 10.1007/978-3-319-39502-9

Library of Congress Control Number: 2016940829

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Dedicated to my wife Inge

Preface

The theory of probability and stochastic processes is often neglected in the education of chemists and biologists, although modern experimental techniques allow for investigations of small sample sizes down to single molecules and provide experimental data that are sufficiently accurate for direct detection of fluctuations. Progress in the development of new techniques and improvement in the resolution of conventional experiments have been enormous over the last 50 years. Indeed, molecular spectroscopy has provided hitherto unimaginable insights into processes at atomic resolution down to time ranges of a hundred attoseconds, whence observations of single particles have become routine, and as a consequence current theory in physics, chemistry, and the life sciences cannot be successful without a deeper understanding of fluctuations and their origins. Sampling of data and reproduction of processes are doomed to produce interpretation artifacts unless the observer has a solid background in the mathematics of probabilities. As a matter of fact, stochastic processes are much closer to observation than deterministic descriptions in modern science, as indeed they are in everyday life, and presently available computer facilities provide new tools that can bring us closer to applications by supplementing analytical work on stochastic phenomena with simulations.

The relevance of fluctuations in the description of real-world phenomena ranges, of course, from unimportant to dominant. The motions of planets and moons as described by celestial mechanics marked the beginning of modeling by means of differential equations. Fluctuations in these cases are so small that they cannot be detected, not even by the most accurate measurements: sunrise, sunset, and solar eclipses are predictable with almost no scatter. Processes in the life sciences are entirely different. A famous and typical historical example is Mendel's laws of inheritance: regularities are detectable only in sufficiently large samples of individual observations, and the influence of stochasticity is ubiquitous. Processes in chemistry lie between the two extremes: the deterministic approach in conventional chemical reaction kinetics has not become less applicable, nor have the results become less reliable in the light of modern experiments. What has increased dramatically are the accessible resolutions in amounts of materials, space, and

time. Deeper insights into mechanisms provide new access to information regarding molecular properties for theory and practice.

Biology is currently in a state of transition: the molecular connections with chemistry have revolutionized the sources of biological data, and this sets the stage for a new theoretical biology. Historically, biology was based almost exclusively on observation and theory in biology engaged only in the interpretation of observed regularities. The development of biochemistry at the end of the nineteenth and the first half of the twentieth century introduced quantitative thinking concerning chemical kinetics into some biological subdisciplines. Biochemistry also brought a new dimension to experiments in biology in the form of *in vitro* studies on isolated and purified biomolecules. A second influx of mathematics into biology came from population genetics, first developed in the 1920s as a new theoretical discipline uniting Darwin's natural selection and Mendelian genetics. This became part of the theoretical approach more than 20 years before evolutionary biologists completed the so-called *synthetic theory*, achieving the same goal.

Then, in the second half of the twentieth century, molecular biology started to build a solid bridge from chemistry to biology, and the enormous progress in experimental techniques created a previously unknown situation in biology. Indeed, the volume of information soon went well beyond the capacities of the human mind, and new procedures were required for data handling, analysis, and interpretation. Today, biological cells and whole organisms have become accessible to complete description at the molecular level. The overwhelming amount of information required for a deeper understanding of biological objects is a consequence of two factors: (i) the complexity of biological entities and (ii) the lack of a universal theoretical biology.

Primarily, apart from elaborate computer techniques, the current flood of results from molecular genetics and genomics to systems biology and synthetic biology requires suitable statistical methods and tools for verification and evaluation of data. However, analysis, interpretation, and understanding of experimental results are impossible without proper modeling tools. In the past, these tools were primarily based on differential equations, but it has been realized within the last two decades that an extension of the available methodological repertoire by stochastic methods and techniques from other mathematical disciplines is inevitable. Moreover, the enormous complexity of the genetic and metabolic networks in the cell calls for radically new methods of modeling that resemble the mesoscopic level of description in solid state physics. In mesoscopic models, the overwhelming and for many purposes dispensable wealth of detailed molecular information is cast into a partially probabilistic description in the spirit of *dissipative particle dynamics* [358, 401], for example, and such a description cannot be successful without a solid mathematical background.

The field of stochastic processes has not been bypassed by the digital revolution. Numerical calculation and computer simulation play a decisive role in present-day stochastic modeling in physics, chemistry, and biology. Speed of computation and digital storage capacities have been growing exponentially since the 1960s, with a doubling time of about 18 months, a fact commonly referred to as Moore's law

[409]. It is not so well known, however, that the spectacular exponential growth in computer power has been overshadowed by progress in numerical methods, as attested by an enormous increase in the efficiency of algorithms. To give just one example, reported by Martin Grötschel from the Konrad Zuse-Zentrum in Berlin [260, p. 71]:

> The solution of a benchmark production planning model by linear programming would have taken – extrapolated – 82 years CPU time in 1988, using the computers and the linear programming algorithms of the day. In 2003 – fifteen years later – the same model could be solved in one minute and this means an improvement by a factor of about 43 million. Out of this, a factor of roughly 1 000 resulted from the increase in processor speed whereas a factor of 43 000 was due to improvement in the algorithms.

There are many other examples of similar progress in the design of algorithms. However, the analysis and design of high-performance numerical methods require a firm background in mathematics. The availability of cheap computing power has also changed the attitude toward exact results in terms of complicated functions: it does not take much more computer time to compute a sophisticated hypergeometric function than to evaluate an ordinary trigonometric expression for an arbitrary argument, and operations on confusingly complicated equations are enormously facilitated by symbolic computation. In this way, present-day computational facilities can have a significant impact on analytical work, too.

In the past, biologists often had mixed feelings about mathematics and reservations about using too much theory. The new developments, however, have changed this situation, if only because the enormous amount of data collected using the new techniques can neither be inspected by human eyes nor comprehended by human brains. Sophisticated software is required for handling and analysis, and modern biologists have come to rely on it [483]. The biologist Sydney Brenner, an early pioneer of molecular life sciences, makes the following point [64]:

> But of course we see the most clear-cut dichotomy between hunters and gatherers in the practice of modern biological research. I was taught in the pregenomic era to be a hunter. I learnt how to identify the wild beasts and how to go out, hunt them down and kill them. We are now, however, being urged to be gatherers, to collect everything lying about and put it into storehouses. Someday, it is assumed, someone will come and sort through the storehouses, discard all the junk and keep the rare finds. The only difficulty is how to recognize them.

The recent developments in molecular biology, genomics, and organismic biology, however, seem to initiate this change in biological thinking, since there is practically no way of shaping modern life sciences without mathematics, computer science, and theory. Brenner advocates the development of a comprehensive theory that would provide a proper framework for modern biology [63]. He and others are calling for a *new theoretical biology* capable of handling the enormous biological complexity. Manfred Eigen stated very clearly what can be expected from such a theory [112, p. xii]:

> Theory cannot remove complexity but it can show what kind of 'regular' behavior can be expected and what experiments have to be done to get a grasp on the irregularities.

Among other things, the new theoretical biology will have to find an appropriate way to combine randomness and deterministic behavior in modeling, and it is safe to predict that it will need a strong anchor in mathematics in order to be successful.

In this monograph, an attempt is made to bring together the mathematical background material that would be needed to understand stochastic processes and their applications in chemistry and biology. In the sense of the version of Occam's razor attributed to Albert Einstein [70, pp. 384–385; p. 475], viz., "everything should be made as simple as possible, but not simpler," dispensable refinements of higher mathematics have been avoided. In particular, an attempt has been made to keep mathematical requirements at the level of an undergraduate mathematics course for scientists, and the monograph is designed to be as self-contained as possible. A reader with sufficient background should be able to find most of the desired explanations in the book itself. Nevertheless, a substantial set of references is given for further reading. Derivations of key equations are given wherever this can be done without unreasonable mathematical effort. The derivations of analytical solutions for selected examples are given in full detail, because readers interested in applying the theory of stochastic processes in a practical context should be in a position to derive new solutions on their own. Some sections that are not required if one is primarily interested in applications are marked by a star ($\star$) for skipping by readers who are willing to accept the basic results without explanations.

The book is divided into five chapters. The first provides an introduction to probability theory and follows in part the introduction to probability theory by Kai Lai Chung [84], while Chap. 2 deals with the link between abstract probabilities and measurable quantities through statistics. Chapter 3 describes stochastic processes and their analysis and has been partly inspired by Crispin Gardiner's handbook [194]. Chapters 4 and 5 present selected applications of stochastic processes to problem-solving in chemistry and biology. Throughout the book, the focus is on stochastic methods, and the scientific origin of the various equations is never discussed, apart from one exception: chemical kinetics. In this case, we present two sections on the theory and empirical determination of reaction rate parameters, because for this example it is possible to show how Ariadne's red thread can guide us from first principles in theoretical physics to the equations of stochastic chemical kinetics. We have refrained from preparing a separate section with exercises, but case studies which may serve as good examples of calculations done by the reader himself are indicated throughout the book. Among others, useful textbooks would be [84, 140, 160, 161, 194, 201, 214, 222, 258, 290, 364, 437, 536, 573]. For a brief and concise introduction, we recommend [277]. Standard textbooks in mathematics used for our courses were [21, 57, 383, 467]. For dynamical systems theory, the monographs [225, 253, 496, 513] are recommended.

This book is derived from the manuscript of a course in stochastic chemical kinetics for graduate students of chemistry and biology given in the years 1999, 2006, 2011, and 2013. Comments by the students of all four courses were very helpful in the preparation of this text and are gratefully acknowledged. All figures in this monograph were drawn with the *COREL* software and numerical computations were done with *Mathematica 9*. *Wikipedia, the free encyclopedia*, has been used

extensively by the author in the preparation of the text, and the indirect help by the numerous contributors submitting entries to Wiki is thankfully acknowledged.

Several colleagues gave important advice and made critical readings of the manuscript, among them Edem Arslan, Reinhard Bürger, Christoph Flamm, Thomas Hoffmann-Ostenhof, Christian Höner zu Siederissen, Ian Laurenzi, Stephen Lyle, Eric Mjolsness, Eberhard Neumann, Paul E. Phillipson, Christian Reidys, Bruce E. Shapiro, Karl Sigmund, and Peter F. Stadler. Many thanks go to all of them.

Wien, Austria Peter Schuster
April 2016

Contents