Springer Theses
Recognizing Outstanding Ph.D. Research

Peter Ashcroft

The Statistical Physics of Fixation and Equilibration in Individual-Based Models

Springer

Springer Theses

Recognizing Outstanding Ph.D. Research

Aims and Scope

The series "Springer Theses" brings together a selection of the very best Ph.D. theses from around the world and across the physical sciences. Nominated and endorsed by two recognized specialists, each published volume has been selected for its scientific excellence and the high impact of its contents for the pertinent field of research. For greater accessibility to non-specialists, the published versions include an extended introduction, as well as a foreword by the student's supervisor explaining the special relevance of the work for the field. As a whole, the series will provide a valuable resource both for newcomers to the research fields described, and for other scientists seeking detailed background information on special questions. Finally, it provides an accredited documentation of the valuable contributions made by today's younger generation of scientists.

Theses are accepted into the series by invited nomination only and must fulfill all of the following criteria

- They must be written in good English.
- The topic should fall within the confines of Chemistry, Physics, Earth Sciences, Engineering and related interdisciplinary fields such as Materials, Nanoscience, Chemical Engineering, Complex Systems and Biophysics.
- The work reported in the thesis must represent a significant scientific advance.
- If the thesis includes previously published material, permission to reproduce this must be gained from the respective copyright holder.
- They must have been examined and passed during the 12 months prior to nomination.
- Each thesis should include a foreword by the supervisor outlining the significance of its content.
- The theses should have a clearly defined structure including an introduction accessible to scientists not expert in that particular field.

More information about this series at http://www.springer.com/series/8790

Peter Ashcroft

The Statistical Physics of Fixation and Equilibration in Individual-Based Models

Doctoral Thesis accepted by
The University of Manchester, UK

Springer

Author
Dr. Peter Ashcroft
Institute for Integrative Biology
ETH Zürich
Zürich
Switzerland

Supervisor
Dr. Tobias Galla
School of Physics and Astronomy
The University of Manchester
Manchester
UK

ISSN 2190-5053 ISSN 2190-5061 (electronic)
Springer Theses
ISBN 978-3-319-41212-2 ISBN 978-3-319-41213-9 (eBook)
DOI 10.1007/978-3-319-41213-9

Library of Congress Control Number: 2016943809

Printed on acid-free paper

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

Parts of this thesis have been published in the following journal articles:

P. Ashcroft, P. M. Altrock, and T. Galla, Fixation in finite populations evolving in fluctuating environments. *J. R. Soc. Interface* **11**, 20140663 (2014).

P. Ashcroft, F. Michor, and T. Galla, Stochastic tunneling and metastable states during the somatic evolution of cancer. *Genetics* **199**, 1213 (2015).

P. Ashcroft, A. Traulsen, and T. Galla, When the mean is not enough: Calculating fixation time distributions in birth–death processes. *Phys. Rev. E* **92**, 042154 (2015).

Supervisor's Foreword

A physicist's view of the world used to be rather deterministic 200 years ago. If the state of all particles and the forces acting between them were known, Laplace hypothesised, a 'being' vast enough to project forward these laws could calculate all future states of the universe. This concept of a Laplacian daemon is of course hopelessly out of date. The discovery of quantum mechanics has fundamentally changed our view of the world; the way we think about determinism and predictability is rather different than at the times of Newton and later Laplace. Quantum physics describes randomness which is (as far as we know) intrinsic to nature, and which cannot be overcome, no matter how hard we try.

Perhaps as important as such aleatoric uncertainty is a second source of randomness in mathematical models of the world around us. Towards the late nineteenth century revolutionary ideas about how one might describe systems with large numbers of interacting particles were developed by heroic figures such as Boltzmann and Gibbs. The theory of 'statistical mechanics' was born. These concepts are founded on the idea that the detailed trajectories of each and every particle in a large system are not all that interesting. Instead it is the behaviour of the system as a whole that we care about. The logical consequence is to study ensembles of particles, and their 'statistics'. Unlike in Newtonian physics we no longer ask: Where is this particular particle going to be at a later time? Instead we ask: If the initial distribution of particles is this, what is the probability to find a given particle in a certain area of space, or with a given speed?

This leads to stochastic descriptions of the laws of physics, the equations governing the dynamics are now subject to noise. This is so-called 'epistemic noise', it originates from the way we model those ensembles of particles. Leaving quantum physics aside, epistemic uncertainty could be eliminated by making a more detailed model and including all forces and interactions in a Laplacian sense. But as statistical physicists we decide not to, because these are not the relevant questions. What is relevant is how global large-scale behaviour emerges from microscopic interaction, not the microscopic trajectories as such.

In order to do this a whole new theory—the theory of stochastic processes—had to be invented. Much of the twentieth century was spent on trying to understand and classify the equilibrium states interacting particle systems reach in the long term. This work is now largely complete, and the focus has moved to systems out of equilibrium. These are systems which do not settle down, they are subject to driving, fluxes, and coupling to the external surroundings. No coherent theory exists for the physics far from equilibrium, but at the same time many pressing challenges rely on progress in this field. This includes turbulence, plasma fusion, active matter, quantum materials, and most notably the physics of life. Biology is inherently out of equilibrium and based on transport of nutrients, energy, the absorption of light, sudden changes, large deviations, the dynamics of evolution and changing environments. It is no surprise that physicists have been able to make remarkable contributions, and that ideas from statistical physics and the theory of stochastic processes have delivered important advances.

Peter joined this adventure in 2012, his thesis focuses on the dynamics of fixation in models of interacting individuals. Peter has investigated several problems at the boundary of theoretical physics and biology. The thesis contains the study of an evolutionary model of populations in switching environments, relevant for example for antibiotic treatment in colonies of bacteria. He has also analysed metastable states in a model of cancer initiation, and the relation of so-called mixing times and the dynamics of fixation in birth–death processes. Chapter 6 of his thesis presents a pedagogical account of the so-called WKB method, a technique from semi-classical physics used to study phenomena including epidemics, ecosystems and, in Peter's thesis, models of cancer evolution.

While I mention the words 'cancer', 'bacteria' and 'evolution' we should be clear: this is a thesis in statistical physics. It contains mathematics and long equations, things are complicated and subtle. The problems Peter looks at are motivated in biology, but the true beauty of this thesis is in the beauty of the underlying mathematical structures and the theoretical concepts and ideas used to unveil them.

Santiago de Compostela, Spain Dr. Tobias Galla
April 2016

Abstract

Individual-based models have been applied to study a broad spectrum of problems across multiple disciplines, such as the spread of epidemics or the outcome of social dilemma. They are used to investigate the macroscopic effects that arise from the microscopic dynamics of interacting individuals. Fixation describes the taking over of the population by a single type of individual or species. It is a prominent feature in the field of population genetics, which interprets many biological scenarios of evolution. Equilibration describes the process of reaching a heterogeneous steady state. In this thesis we analyse these macroscopic features through techniques derived from statistical physics and the theory of stochastic processes.

Birth–death processes are used to describe the interaction of two types of individual in a population, such as competing strains of bacteria. These interactions are often specified using the framework of evolutionary game theory. The environment in which the population evolves can have a crucial impact on selection. In systems where the environment switches between multiple states we develop a general theory to calculate the fixation time statistics of a mutant individual in a population of wild-types, as well as the stationary distributions when mutations are present in the dynamics. In some birth–death processes, and in particular those described by evolutionary game theory, the mean fixation time contains only limited information. By diagonalising the master equation that describes the process, we are able to obtain closed-form expressions for the complete fixation time distributions.

Individual-based models can also be used to describe the accumulation of mutations in a cell. This has important consequences for the initiation and progression of cancer. We find that such systems exhibit metastable states in the dynamics, and we can exploit the separation of timescales between relaxing to the quasi-stationary state and reaching fixation to characterise these phenomena. In this scenario we employ the WKB method to describe the population-level dynamics. Although this method has been used to describe numerous stochastic processes, a clear and coherent description is lacking in the literature. Through the use of multiple examples, including the aforementioned cancer initiation model, we carefully explain the multitude of constructs and equations that result from the application of this method.

The analytical characterisation of the evolutionary dynamics that are observed in these stochastic processes has resulted in a greater understanding of fixation and equilibration. This thesis promotes the benefits of analytical, or even semi-analytical methods, and on a more general level contributes towards a more complete understanding of evolutionary processes.

Acknowledgments

I would first like to thank my supervisor Tobias Galla, whose infectious enthusiasm and invaluable guidance have ensured this experience has been wholly fulfilling. My gratitudes are also with Alan McKane for maintaining a stimulating environment for me to work in, and with my collaborators Philipp Altrock, Franziska Michor, and Arne Traulsen for the support they provided and the opportunities they presented. A special mention is reserved for Yen Ting Lin for his much-appreciated feedback and constructive criticisms.

Thanks to all the occupants of office 7.26 who are too numerous to name, but you know who you are, and to the postdocs who have been and gone during my Ph.D.

Thanks to my niece Mia, whose timely birth in the middle of writing this thesis provided some much-needed distraction, and to the rest of my family for their support.

Finally, I want to give a really big thank you to Stacey, who has consistently put up with my unending drivel about the wonders of science. Without your support and understanding I would not be who I am today. Here's to the next adventure!

Contents

About the Author

Dr. Peter Ashcroft graduated as a Master of Mathematics and Physics from the University of Manchester in 2012. He then studied for his Ph.D. in theoretical physics under the supervision of Dr. Tobias Galla in Manchester. During his Ph.D., Peter was especially interested in the theory of stochastic processes and in the mathematical modelling of biological phenomena. He completed his Ph.D. in 2015, and currently is a postdoc in theoretical biology at ETH Zürich where, amongst other projects, he is investigating the dynamics of blood formation and disease.

Chapter 1
Introduction

Over the next 200 or so pages I will explore how tools and concepts developed within theoretical physics can be applied to problems in other sciences. Although more emphasis in this thesis will be directed to biological applications, the successes of this field can also be seen in social science [1, 2], economics [3, 4], and many other disciplines where so-called complex systems are a prominent feature.

My motivation for working in this area is the freedom you have to explore these numerous disciplines, and the satisfaction that arises from solving a long-standing problem by approaching it from an unconventional point of view. Interactions with academics from these various backgrounds has provided hours of intellectual conversation and brainstorming that have greatly enhanced my knowledge of the world outside of physics. But ultimately the main reason for joining this area, and choosing to continue my career in this field, is because the analysis is fun! The benefits of the approaches I use lie not only in their predictive power, but they are enjoyable, satisfying, stimulating and infuriating in equal measures.

The success of theoretical physics across multiple disciplines comes from its ability to break down objects to their fundamental constituents. Analysis of the inner workings then allows the practitioner to obtain a more complete understanding of the world. An experimentalist works with the real-world system, or a synthetic *in vitro* analogue. Their understanding of this system is achieved through the collection and analysis of data. Theorists, however, obtain an understanding by considering a representation of the real-world system, which I will refer to as a model.

For biological systems an exact model representation is often impossible due to the inherent complexity of many interacting entities. If a model is almost as complicated as the experimental system, it will be just as intractable. In the end you would have the same data set, but generated *in silico*, and no new insight or understanding will have been gained. As the level of abstraction from the real world increases, so does the level of tractability. The balance between accuracy and tractability is a choice to be made by the modeller. In the case of this thesis, Occam's Razor prevails;

P. Ashcroft, *The Statistical Physics of Fixation and Equilibration in Individual-Based Models*, Springer Theses, DOI 10.1007/978-3-319-41213-9_1

I will focus on the simplest models which reproduce observed behaviours, but can be applied to a wide range of problems. These models can highlight the underlying mechanisms that result in the observed phenomenon, something that may not be immediately obvious from simply conducting an experiment.

One of the most profound examples of this in the biosciences is the explanation of the regular structures on the coats of animals [5]. The colouration was known to be caused by melanin in the skin, but there was no explanation for the origin of the pattern of this colouration in animals such as zebras and leopards. The seminal work of Alan Turing (1912–1954) provided part of the answer. Turing proposed that diffusive chemicals can settle into a stable, spatially-inhomogeneous state through the excitation of the now-called Turing instability [6]. Although the true mechanism is more complex than the idea proposed by Turing [7], the same basic principles were applied to reproduce observed animal coat patterns [8].

The class of systems in which my interest lies is not the continuous reaction–diffusion systems as studied by Turing, but systems that contain a finite number of discrete, interacting 'particles' or individuals. Such systems are ubiquitous in nature, where particles could represent proteins, molecules, cells, bacteria, animals or people. The dynamics of the particles can be governed by events such as production (birth), degradation (death), predation or infection, to name but a few. The discreteness of the particles, and the nature of the dynamics, are responsible for the observed stochasticity; that is, there is an intrinsic source of randomness in these systems, often referred to as demographic noise.

The discreteness of the particles, and with it the intrinsic stochasticity, is retained when modelling these systems. However, information about the behaviour of every individual particle is not necessary. Instead, the simplifying assumption that two particles of the same type are indistinguishable is made. The behaviour of the system can then be described by the statistics of the group of particles. This procedure is the basis of statistical mechanics, and the approach is poetically summarised by James Clerk Maxwell (1831–1879):

> And here I wish to point out that, in adopting this statistical method of considering the average number of groups of molecules selected according to their velocities, we have abandoned the strict kinetic method of tracing the exact circumstances of each individual molecule in all its encounters.
>
> It is therefore possible that we may arrive at results which, though they fairly represent the facts as long as we are supposed to deal with a gas in mass, would cease to be applicable if our faculties and instruments were so sharpened that we could detect and lay hold of each molecule and trace it through all its course.
>
> James Clerk Maxwell, *The Theory of Heat* [9].

Here Maxwell is referring to the original derivation of the Maxwell–Boltzmann distribution, which describes the distribution of speeds of molecules of a contained ideal gas [10].[1]

[1] Ludwig Boltzmann (1844–1906) later derived this result from the kinetic theory of gases [11].

To model the particles in the discrete systems *in mass* they are treated like molecules of a gas. The interactions then take a form which is similar to that of chemical reactions. These reactions are dependent on the number of reactants (molecules) available and the rate at which the gas molecules interact [12]. These models are referred to as individual-based models, and they have been applied to study epidemic outbreaks [13], social dilemma [14], predator–prey interaction [15], and the list can go on and on. This thesis, however, is not dedicated to a particular system or application. Instead I will investigate particular phenomena that are observed in a variety of stochastic systems. These are:

Fixation: The process of a single type of individual taking over the whole population. The term originates from the field of population genetics, where the fixation of an allele was a central topic [16–19]. In this case fixation occurs when all other alleles are irreversibly lost from the gene pool, and only a single *fixed* allele remains. The terminology is now used outside of population genetics and the study of gene frequencies, for example to describe the eradication of a disease or reaching a social consensus.

Equilibration: The process of reaching a stable stationary state. If fixation is not possible in a system, as is the case if individuals can change their type stochastically, then the success of a type of individual is no longer characterised by the probability that it takes over the population. Instead success can be measured by its relative concentration at long times. This is described by the stationary probability distribution. The time to approach this stable state is also of interest.

These two effects are closely linked; if a system fixates then no more dynamics can occur and hence the fixated state is stationary. They are also related if fixation takes a very long time, such that the system can initially relax into a quasi-stationary state before fixation occurs. These links will be investigated closely in Chaps. 4 and 5.

A concrete understanding of the effects of fixation and equilibration, and the interplay between them, will greatly contribute to our understanding of the process of evolution. This field of investigating evolution through mathematical approaches has been dubbed evolutionary dynamics, and it describes the change of populations over time subject to spontaneous mutation, selection, and random events [14, 20]. Different types of individual in the population, which we will sometimes call phenotypes in line with the biological literature, can emerge spontaneously by mutation, i.e. through errors during reproduction of the pre-existing *wild-types*. In many cases, wild-type and mutant individuals are characterised by heritable differences in behavioural traits or strategies [14]. Selection acts on different (pheno)types and their associated traits to change the population composition.

One of the great successes of evolutionary dynamics is the quantitative analysis of cancer, which is a genetic disease and according to Cancer Research UK, "*1 in 2 people in the UK born after 1960 will be diagnosed with some form of cancer during their lifetime"* [21]. Mathematical investigations have contributed profoundly to our understanding of "*the emperor of all maladies*" [22]. Numerous studies throughout the 20th century have addressed the kinetics of cancer initiation and progression

[23–28]. In Ref. [23], it was first proposed that "*several successive mutations in the same cell […] would be necessary [for cancer to initiate]*". Empirical observations of mortality rates across a range of cancer types agreed with this hypothesis [24]. For some varieties of cancer it was shown by Alfred Knudson (1922–) that tumours can be induced by as few as two mutations, corresponding to the inactivation of both copies of a specific tumour suppressor gene (TSG) [26]. The data that confirmed this hypothesis is presented in Fig. 1.1. This is data for the diagnosis of tumours, or retinoblastomas, in the eyes of children. Knudson hypothesised that if the tumours required two mutations, we would observe a quadratic incidence rate. However, if the child had inherited a defective gene, the incidence curve should be linear and there is a much larger probability that the cancer will be present in both eyes, which is referred to as bilateral. The data clearly favours Knudson's interpretation, and this is the celebrated *two-hit hypothesis* [26].

The age of stochastic modelling of cancer initiation began in earnest with the introduction of the branching process, as shown in Fig. 1.2 [28]. Similar models have been used extensively to describe various aspects of carcinogenesis [29, 30],

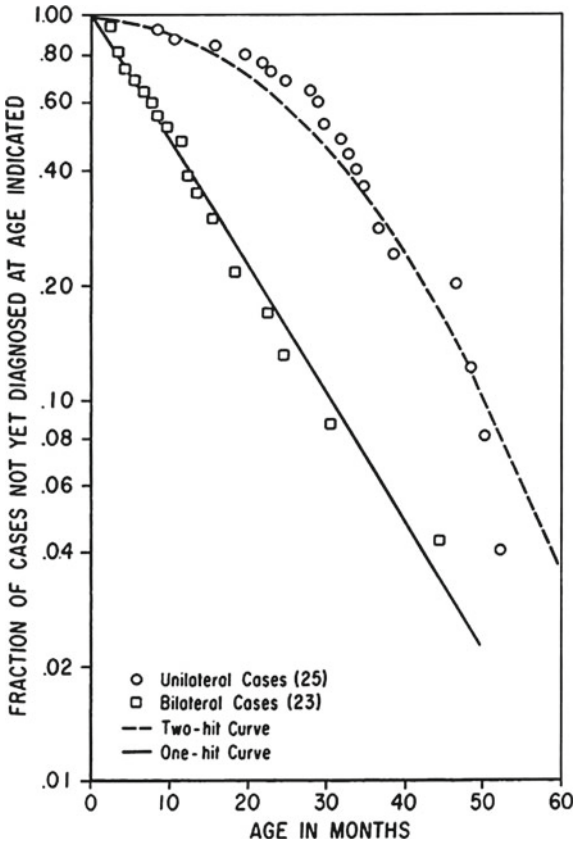

Fig. 1.1 Fraction of cases of retinoblastoma not yet diagnosed as a function of the children's age. The one-hit (bilateral) curve is $\log S = -t/30$, and the two-hit (unilateral) curve is $\log S = 4 \times 10^{-5} t^2$, where S is the fraction of cases not diagnosed and t is the children's age in months. This figure is from Ref. [26]