Interdisciplinary Applied Mathematics 41

Paul C. Bressloff

Stochastic Processes in Cell Biology

Springer

Stochastic Processes in Cell Biology

Interdisciplinary Applied Mathematics

Problems in engineering, computational science, and the physical and biological sciences are using increasingly sophisticated mathematical techniques. Thus, the bridge between the mathematical sciences and other disciplines is heavily traveled. The correspondingly increased dialog between the disciplines has led to the establishment of the series: *Interdisciplinary Applied Mathematics*.

The purpose of this series is to meet the current and future needs for the interaction between various science and technology areas on the one hand and mathematics on the other. This is done, firstly, by encouraging the ways that mathematics may be applied in traditional areas, as well as point towards new and innovative areas of applications; and, secondly, by encouraging other scientific disciplines to engage in a dialog with mathematicians outlining their problems to both access new methods and suggest innovative developments within mathematics itself.

The series will consist of monographs and high-level texts from researchers working on the interplay between mathematics and other fields of science and technology.

For further volumes:
http://www.springer.com/series/1390

Paul C. Bressloff

Stochastic Processes in Cell Biology

Springer

Paul C. Bressloff
Department of Mathematics
University of Utah
Salt Lake City, UT, USA

ISSN 0939-6047 ISSN 2196-9973 (electronic)
ISBN 978-3-319-08487-9 ISBN 978-3-319-08488-6 (eBook)
DOI 10.1007/978-3-319-08488-6
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014945210

Mathematics Subject Classification: 82C31, 92C05, 92C37, 92C40, 92C17, 82C70

Printed on acid-free paper

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

To my wife Alessandra

Preface

In recent years there has been an explosion of interest in the effects of noise in cell biology. This has partly been driven by rapid advances in experimental techniques, including high-resolution imaging and molecular-level probes. However, it is also driven by fundamental questions raised by the ubiquity of noise. For example, how does noise at the molecular and cellular levels translate into reliable or robust behavior at the macroscopic level? How do microscopic organisms detect weak environmental signals in the presence of noise? Have single-cell and more complex organisms evolved to exploit noise to enhance performance? In light of the above, there is a growing need for mathematical biologists and other applied mathematicians interested in biological problems to have some background in applied probability theory and stochastic processes. Traditional mathematical courses and textbooks in cell biology and cell physiology tend to focus on deterministic models based on differential equations such as the Hodgkin–Huxley and FitzHugh–Nagumo equations, chemical kinetic equations, and reaction–diffusion equations. Although there are a number of well-known textbooks on applied stochastic processes, they are written primarily for physicists and chemists or for population biologists. There are also several excellent books on cell biology written from a biophysics perspective. However, these assume some background in statistical physics and a certain level of physical intuition. Therefore, I felt that it was timely to write a textbook for applied mathematicians interested in learning stochastic processes within the context of cell biology, which could also serve as an introduction to mathematical cell biology for statistical physicists and applied probabilists.

I started my interest in stochastic cell biology, as distinct from my work in mathematical neuroscience, around 8 years ago when I volunteered to teach a course in biophysics for the mathematical biology graduate program at Utah. I was immediately fascinated by the molecular processes underlying the operation of a cell, particularly the mechanisms for transporting proteins and other macromolecules to the correct subcellular targets at the correct times. Such an issue is particularly acute for neurons, which are among the largest and most complex cells in biology. In healthy cells, the regulation of protein trafficking within a neuron provides an important mechanism for modifying the strength of synaptic connections between neurons,

and synaptic plasticity is generally believed to be the cellular substrate of learning and memory. On the other hand, various types of dysfunction in protein trafficking appear to be a major contributory factor to a number of neurodegenerative diseases associated with memory loss including Alzheimer's disease.

In writing this book, I have gone back to my roots in theoretical physics, but refracted through the lens formed by many years of working in applied mathematics. Hence, the book provides extensive coverage of analytical methods such as initial boundary value problems for partial differential equations, singular perturbation theory, slow/fast analysis and quasi-steady-state approximations, Green's functions, WKB methods and Hamilton–Jacobi equations, homogenization theory and multi-scale analysis, the method of characteristics and shocks, and reaction–diffusion equations. I have also endeavored to minimize the use of statistical mechanics, which is not usually part of a mathematician's tool-kit and requires a certain level of physical intuition. It is not possible to avoid this topic completely, since many experimental and theoretical papers in cell biology assume some familiarity with terms such as entropy, free energy, and chemical potential. The reason is that microscopic systems often operate close to thermodynamic equilibrium or asymptotically approach thermodynamic equilibrium in the long-time limit. This then imposes constraints on any model of the underlying stochastic process. In most cases, one can understand these constraints by considering the Boltzmann–Gibbs distribution of a macromolecule in thermodynamic equilibrium, which is the approach I take in this book.

There are two complementary approaches to modeling biological systems. One involves a high level of biological detail and computational complexity, which means that it is usually less amenable to mathematical analysis than simpler reduced models. The focus tends to be on issues such as parameter searches and data fitting, sensitivity analysis, model reductions, numerical convergence, and computational efficiency. This is exemplified by the rapidly growing field of systems biology. The other approach is based on relatively simple conceptual or "toy" models, which are analytically tractable and, hopefully, capture essential features of the phenomena of interest. In this book I focus on the latter for pedagogical reasons and because of my own personal tastes. In the introductory chapter, I summarize some of the basic concepts in stochastic processes and nonequilibrium systems that are used throughout the book, describe various experimental methods for probing noise at the molecular and cellular levels, give a brief review of basic probability theory and statistical mechanics, and then highlight the structure of the book. In brief, the book is divided into two parts: Part I (Foundations) and Part II (Advanced Topics). Part I provides the basic foundations of both discrete and continuous stochastic processes in cell biology. Its five chapters deal with diffusion, random walks, and the Fokker–Planck equation (Chap. 2), stochastic ion channels (Chap. 3), polymers and molecular motors (Chap. 4), biochemical signaling and adaptation (Chap. 5), and gene expression and regulatory networks (Chap. 6). Part II covers more advanced topics that build upon the ideas and techniques from Part I. Topics include transport processes in cells (Chap. 7), self-organization of the cytoskeleton (Chap. 8), self-organization in reaction–diffusion models (Chap. 9), WKB methods

for escape problems (Chap. 10), and some more advanced topics in probability theory (Chap. 11). The chapters are supplemented by additional background material highlighted in gray boxes and numerous exercises that reinforce the analytical methods and models introduced in the main body of the text. I have attempted to make the book as self-contained as possible. However, some introductory background in partial differential equations, integral transforms, and applied probability theory would be advantageous.

Finally, this book should come with a "government health warning." That is, throughout most of the book, I review the simplest mechanistic models that have been constructed in order to investigate a particular biological phenomenon or illustrate a particular mathematical method. Although I try to make clear the assumptions underlying each model, I do not carry out a comparative study of different models in terms of the degree of quantitative agreement with experimental data. Therefore, the reader should be cautioned that the models are far from the last word on a given phenomenon, and the real biological system is usually way more complicated than stated. However, it is hoped that the range of modeling and analytical techniques presented in this book, when combined with efficient numerical methods, provides the foundations for developing more realistic, quantitative models in stochastic cell biology.

Salt Lake City, UT, USA Paul C. Bressloff

Supplementary material - available online:

Springer.com web page for the book:
www.springer.com/978-3-319-08487-9 where there will be a password protected solution manual in the near future.

Author website for supplementary material (updates, additional exercises, further topics):
http://www.math.utah.edu/~bresslof/#Books

Acknowledgments Much of my recent research on stochastic processes in cell biology has been done in a very fruitful and enjoyable collaboration with my ex-graduate student Jay Newby. There are many applied mathematicians, physical scientists, and life scientists upon whose sturdy shoulders I have stood during the writing of this book and whose work is featured extensively in the following pages. I apologize in advance if I have excluded anyone or didn't do proper justice to their contributions. Finally, I would like to thank my wife Alessandra and son Luca (the Shmu) for their love and support, and for putting up with my hermit-like existence over the past year.

Contents

List of Boxes

Chapter 1
Introduction

One of the major challenges in modern biology is to understand how the molecular components of a living cell operate in a highly noisy environment. What are the specific sources of noise in a cell? How do cells attenuate the effects of noise in order to exhibit reliable behavior (robustness to noise)? In particular, how does a stochastic genotype result in a reliable phenotype through development? How does the noisy, crowded environment of a cell affect diffusive transport? How do molecular machines convert chemical energy to work? What are the physical limits of biochemical signaling, such as the sensitivity of biochemical sensors to environmental signals? Under what circumstances can a cell exploit noise to enhance its performance or the survival of its host organism? What is the role of self-organization in the formation and maintenance of subcellular structures such as the cytoskeleton? The goal of this book is to use the theory of stochastic processes and nonequilibrium systems to investigate these types of biological questions at the cellular level; analogous questions also hold at the multicellular level but are not addressed in this book since they would double its length! One can view the book either as an introduction to stochastic processes using cell biology as the motivating application or, conversely, as an introduction to mathematical cell biology with an emphasis on stochastic processes. Irrespective of the particular perspective, it is clear that there is a growing demand for mathematical biologists and other applied mathematicians to have some training in topics that have traditionally been the purview of physicists, chemists, and probabilists. This book provides the necessary background to tackle research problems in mathematical biology that involve stochastic processes.

1.1 Stochastic Processes in Living Cells

In this first section, we introduce some of the basic concepts that are useful in characterizing and analyzing noise in cells, starting at the level of individual macromolecules and building up to cellular structures. For an excellent general introduction to molecular and cell biology, we refer the reader to the book *Molecular Biology*

P.C. Bressloff, *Stochastic Processes in Cell Biology*, Interdisciplinary Applied Mathematics 41, DOI 10.1007/978-3-319-08488-6_1

of the Cell by Alberts et al. [3]. For a more biophysics-oriented approach to cell biology, see *Physical Biology of the Cell* by Phillips et al. [509] and *Biophysics* by Bialek [44]. An extensive coverage of cell physiology with some examples of stochastic processes can be found in the first volume of *Mathematical Physiology* by Keener and Sneyd. Two standard references on the theory of stochastic processes, with an emphasis on physical and chemical processes, are *Handbook of Stochastic Methods* by Gardiner [204] and *Stochastic Processes in Physics and Chemistry* by Van Kampen [651]. For a more kinetic-based treatment of nonequilibrium processes, see *A Kinetic View of Statistical Physics* by Krapivsky et al. [354]. A book on stochastic processes oriented towards population biology is *An Introduction to Stochastic Processes with Applications to Biology* by Allen [4]. For a more mathematical formulation of discrete stochastic processes, see *Probability and Random Processes* by Grimmett and Strirzaker [242], and for a more rigorous treatment of stochastic differential equations (SDEs), see *SDEs* by Oksendal [483]. Finally, an analytical treatment of stochastic processes with a detailed description of asymptotic methods and large deviations can be found in the book *Theory and Applications of Stochastic Processes* by Schuss [574].

1.1.1 Internal and External States of a Macromolecule

Consider a single macromolecule such as a motor protein, an enzyme, an ion channel, or a strand of DNA. Each macromolecule is subject to thermal fluctuations arising from the continual bombardment by molecules in the surrounding environment, which could be the interior aqueous solution of the cell (cytosol) or the surrounding plasma membrane. The size of molecular fluctuations is set by the basic unit of thermal energy k_BT, where T is the temperature (in degrees Kelvin K) and $k_B \approx 1.4 \times 10^{-23}\text{JK}^{-1}$ is the Boltzmann constant. A useful distinction at the molecular level is between *internal conformational states* of a macromolecule and *external states* such as the position and momentum of the center of mass of the molecule. Often the internal degrees of freedom are represented as a set of discrete states, and the stochastic dynamics within this state space is described in terms of a *continuous-time Markov process* [204]. That is, the state of the system takes values in some finite or countable set, and the time spent in each state has an exponential distribution. Moreover, the continuous-time stochastic process has the Markov property, which means that the future behavior of the system, both the remaining time in the current state and the identity of the next state, depends only on the current state and not on any prior history. A simple two-state continuous-time Markov process can be used to model the opening and closing of an ion channel, for example, or the binding and unbinding of a ligand molecule to a protein receptor (see Fig. 1.1a). More generally, suppose that a macromolecule has m internal states labeled $j = 1, \ldots, m$. The probability $P_j(t)$ that the molecule is in state j at time t evolves according to the system of differential equations

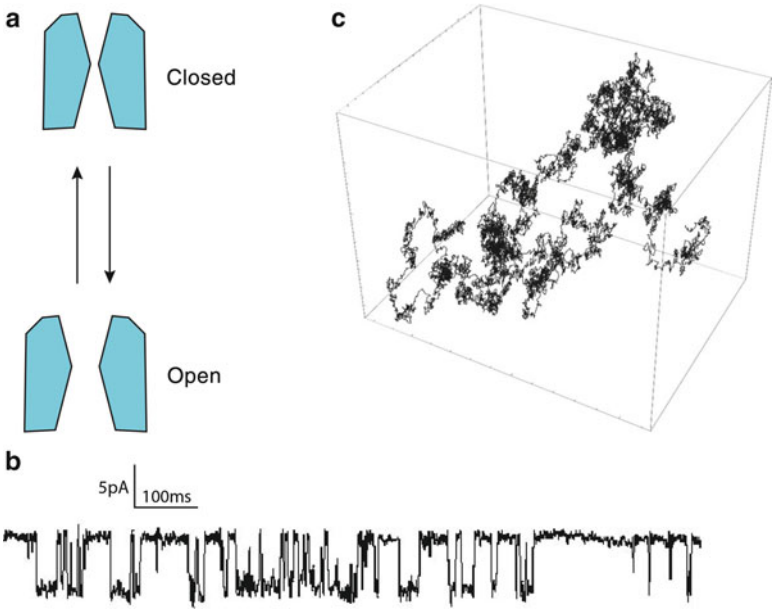


Fig. 1.1: Internal and external stochastic variables. (**a**) Internal open and closed states of an ion channel. (**b**) Patch-clamp recording of a glycine receptor showing stochastic variations in current due to the opening and closing of the ion channel. (Public domain figure downloaded from Wikimedia Commons.) (**c**) Sample 3D trajectory of a Brownian particle

$$\frac{dP_j}{dt} = \sum_{k=1}^{m} \left[W_{jk} P_k(t) - W_{kj} P_j(t) \right], \quad j = 1, \ldots, m,$$

where $W_{jk}\delta t$ is the probability that the molecule jumps to state j in an infinitesimal time interval δt, given that it is currently in state k; W_{jk} is called a state transition rate. Such a Markov process is said to satisfy *detailed balance* if there exists a stationary density Π_j such that for each pair of reversible transitions (jk)

$$W_{jk}\Pi_k = W_{kj}\Pi_j.$$

The detailed balance condition is stronger than that required merely for a stationary distribution—there are Markov processes with stationary distributions that do not have detailed balance. Detailed balance implies that, around any closed cycle of states, there is no net flow of probability.

In contrast to discrete internal states, the evolution of external variables such as the position of a macromolecule is modeled in terms of an SDE [204, 483], which is often called a *Langevin equation* in the physics literature [651]. Mathematically speaking, an SDE is a differential equation in which one or more of the terms are a stochastic process, resulting in a solution that is itself a stochastic process.

In the case of a macromolecule, the stochastic terms represent the effective forces due to collisions of the macromolecule with molecules in the surrounding medium. From this perspective, the external state of the macromolecule is described in terms of slow variables, whereas the degrees of freedom of the surrounding medium are treated as fast microscopic variables that are responsible for the stochastic nature of the SDE. A classical example of an SDE is Brownian motion, which refers to the random motion of a particle suspended in a fluid (see Fig. 1.1c). This phenomenon is named after the botanist Robert Brown, who observed the erratic motion of pollen grains suspended in water through a microscope, but was unable to determine the mechanisms that caused this motion. Albert Einstein subsequently explained how the motion that Brown had observed was a result of the pollen being moved by individual water molecules, which served as a definitive confirmation that atoms and molecules actually exist. An idealized mathematical representation of Brownian motion is the *Wiener process*, whose evolution can be described in terms of infinitesimal increments that are independent random variables generated from a zero mean Gaussian distribution whose variance scales as Δt, where Δt is the time step. One of the characteristic features of a Wiener process is that in the limit $\Delta t \to 0$, the time-dependent solution is continuous, but its time derivative is everywhere infinite, reflecting the absence of a finite time scale. This means that the usual rules of calculus break down and indicates that a Wiener process is an idealization of an actual random physical process, which always has a finite intrinsic time scale such as the time constant of second-order statistical correlations. Another important feature of a Wiener process is that it is the single-particle realization of diffusion. This can be seen from two perspectives. First, each realization of a Wiener process determines a sample trajectory through state space. The associated probability density of sample paths is the solution to a deterministic partial differential equation (PDE) known as a *Fokker–Planck equation* (FPE). In the case of idealized Brownian motion, the FPE is formally identical to the classical diffusion equation. Indeed, if one were to consider a large number N of noninteracting Brownian particles, then, in the large-N limit, the concentration of particles evolves deterministically according to the diffusion equation. Second, if one were to discretize time and space, then a sample trajectory of a Wiener process reduces to an unbiased random walk, which is well known to be a discrete realization of a diffusing particle.

More generally, consider a molecule of mass m moving in one dimension under the influence of an external force $F(x)$. In the absence of thermal fluctuations, the position of the molecule satisfies Newton's law of motion

$$m\frac{d^2x}{dt^2} + \gamma\frac{dx}{dt} = F(x),$$

where γ is a damping or drag coefficient. It turns out that at the microscopic length and velocity scales of molecular dynamics, the aqueous environment of a cell is highly viscous so that inertial terms can be ignored—the particle rapidly reaches the terminal velocity $F(x)/\gamma$. Under such circumstances, when the force due to thermal fluctuations is modeled as a Wiener process, the FPE equation for the probability density $p(x,t)$ takes the form

$$\frac{\partial p}{\partial t} = -\frac{1}{\gamma}\frac{\partial}{\partial x}(F(x)p(x,t)) + D\frac{\partial^2 p(x,t)}{\partial x^2},$$

where D determines the level of noise. Moreover, since the molecules responsible for the fluctuating force are also responsible for the dissipation or damping, one finds that D and γ are related according to the so-called Einstein relation

$$D\gamma = k_B T.$$

This is a specific example of what is known as a *fluctuation–dissipation theorem*, since it relates the rate of relaxation to equilibrium to the size of thermal fluctuations.

A number of important cellular processes at the macromolecular level involve a coupling between continuous external variables and discrete internal variables, which is modeled using a *stochastic hybrid system*. Consider, for example, molecular motors, which are proteins that convert chemical energy into mechanical work [275]. A motor protein undergoes a cyclic sequence of conformational changes after reacting with one or more molecules of a chemical such as adenosine triphosphate (ATP), resulting in the release of chemical energy. This allows the motor to perform work by exerting a force conjugate to a given external variable, e.g., pulling a load while moving along a protein filament—active intracellular transport. Another example of a stochastic hybrid system is a voltage-gated or ligand-gated ion channel, in which the opening and closing of the channel depends on an external variable such as membrane voltage or calcium concentration [322]. Moreover, the dynamics of the given external variable itself depends on the internal state of the ion channel or, more precisely, a population of ion channels.

1.1.2 Equilibrium and Nonequilibrium Systems

One of the fundamental features of a living cell is that it is an open system, i.e., it interacts with the surrounding environment through the exchange of energy and matter (see Fig. 1.2). Moreover, a cell is maintained out of thermodynamic equilibrium, which means that there are nonzero fluxes of energy, matter, and charge flowing between the interior and exterior of the cell. In such cases, one has to model the stochastic dynamics using a continuous-time Markov process or an SDE, for example. Nevertheless, it is sometimes possible to approximate a subcellular system such as an individual macromolecule as being in thermodynamic equilibrium, provided that the rate of relaxation to local equilibrium is fast compared to other processes of interest. (Note, however, that a macromolecule such as a motor protein can only operate if it is maintained out of thermodynamic equilibrium.) In such cases one can exploit the powerful machinery of equilibrium statistical mechanics [102, 317]. In particular, one can make use of the *Boltzmann–Gibbs distribution*. Suppose that a macromolecule has a set of intrinsic states labeled by j and let E_j denote the (free) energy of the molecule in the jth state. Furthermore, assume that the surrounding cellular environment maintains a constant temperature T. A basic

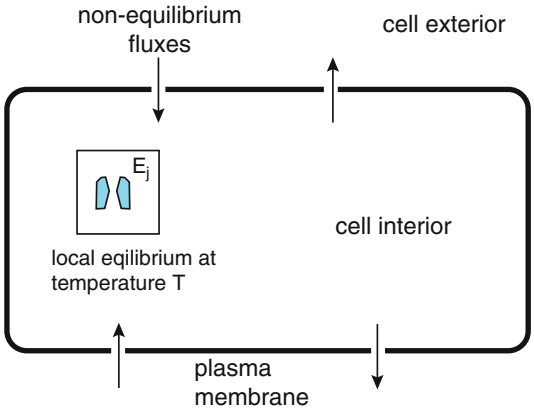


Fig. 1.2: The cell is an open, nonequilibrium system that exchanges energy and matter with the environment. However, a local subsystem such as an individual macromolecule may be in local thermodynamic equilibrium at a temperature T. The probability of being in intrinsic state j with energy E_j is then determined by the Boltzmann–Gibbs distribution

result of equilibrium statistical mechanics is that the probability p_j that the molecule is in state j is given by the Boltzmann–Gibbs distribution (see Sect. 1.4):

$$p_j = \frac{1}{Z} \mathrm{e}^{-E_j/k_B T}, \quad Z = \sum_j \mathrm{e}^{-E_j/k_B T}.$$

The so-called partition function Z ensures that $\sum_j p_j = 1$. If we apply the Boltzmann–Gibbs distribution to the steady-state probability Π_j of detailed balance, we immediately see that the backward and forward transition rates satisfy the condition

$$\frac{W_{jk}}{W_{kj}} = \frac{\Pi_j}{\Pi_k} = \frac{p_j}{p_k} = \mathrm{e}^{-(E_j - E_k)/k_B T}.$$

That is, the ratio of the forward and backward transition rates depends on the energy difference between the two states.

1.1.3 Populations of Molecules

Another important issue is how noise at the level of an individual molecule scales up when one considers a population of possibly interacting molecules. If the population is in thermodynamic equilibrium, then the Boltzmann–Gibbs distribution can still be used, except that now E_j is the total energy of the population in a given microstate, which includes any contributions from molecular interactions within the population. In the case of a large number of molecules, one can often describe the

collective behavior of the population in terms of a small number of macroscopic variables for which thermal fluctuations can be neglected. These macroscopic variables characterize the average or mean-field behavior of the population. Examples include the concentration of molecules within a cellular compartment or the pressure exerted by molecules on the cell membrane. It is also possible to construct a mean-field theory for a population of N independent, identical macromolecules operating out of thermodynamic equilibrium, with each described by a continuous-time Markov process. That is, in the limit $N \to \infty$, the fraction of the population in a given state evolves according to a system of deterministic differential equations (*kinetic equations*). Moreover, for finite N, one can track the stochastic fraction of molecules in a given state using a so-called master equation. Let $\mathbf{n} = (n_1, n_2, \ldots, n_m)$ denote the number of molecules in each of m internal states with $\sum_{j=1}^{m} n_j = N$. The probability that the population is in the configuration state $\mathbf{n}$ at time t then evolves according to a master equation of the form

$$\frac{dP_{\mathbf{n}}}{dt} = \sum_{\mathbf{n}'} \left[W_{\mathbf{n}\mathbf{n}'} P_{\mathbf{n}'}(t) - W_{\mathbf{n}'\mathbf{n}} P_{\mathbf{n}}(t) \right].$$

Although it is generally difficult to analyze such a master equation, it is possible to carry out a perturbation expansion of the master equation in terms of the system size $1/N$ (*system-size expansion*). For large but finite N, one can thus approximate the stochastic dynamics of the population using a FPE or its equivalent Langevin equation [651]. However, now the stochastic variables are the fraction of molecules in a given intrinsic state rather than the position of a single macromolecule undergoing Brownian motion. The FPE then provides an estimate for the size of fluctuations about the mean-field solutions.

In recent years, the system-size expansion of master equations has become a major focus of work on genetic and other biochemical networks within a cell [312, 408, 502, 521, 555]. In the case of classical chemical reactions, the number of molecules involved is huge (comparable to Avagadro's number 6×10^{23}). In such cases, it is sufficient to model the chemical reactions in terms of deterministic kinetic equations based on the *law of mass action*: the rate of an elementary reaction (a reaction that proceeds through only one step) is proportional to the product of the concentrations of the participating molecules. In thermodynamic equilibrium, the rates of the forward and backward reactions must be equal, which allows one to express the ratio of the concentrations of reactants and products in terms of a constant known as the dissociation constant K_d. An expression for K_d can be derived from first principles using the Boltzmann–Gibbs distribution. However, the absolute values of the transition rates (rather than their ratios) cannot be determined from the theory of equilibrium systems. In contrast to classical chemical reactions, the number of molecular constituents involved in the transcription of DNA to produce a protein is small (tens or hundreds of molecules). The low copy numbers mean that fluctuations in the number of proteins produced are non-negligible and one has to deal with the corresponding master equation. An immediate issue that stems from

this is how noisy processes at the level of DNA (genotypes) result in the robust development of single-cell and multicellular organisms (phenotypes).

Once one considers noise at the level of more complex subcellular processes that involve populations of reacting macromolecules, it is useful to distinguish between *intrinsic* and *extrinsic* noise [164, 625]. Intrinsic noise refers to fluctuations due to the inherent stochasticity of the macromolecules, whereas extrinsic noise refers to fluctuations in the external environment (beyond the random molecular collisions that generate the intrinsic noise). For example, given a population of ion channels, intrinsic noise might correspond to fluctuations in the fraction of open ion channels whereas extrinsic noise could be due to random variations in the membrane voltage or calcium concentration. In the case of gene expression, intrinsic noise might refer to fluctuations in the number of bound protein promoters that repress or activate gene expression, whereas extrinsic noise could be due to fluctuations in the rates of binding and unbinding.

1.1.4 Self-Organization

A significant concept in the theory of nonequilibrium systems is the notion of *self-organization*. Within the context of cell biology, this plays an important role in the formation and maintenance of cellular and subcellular structures such as those shown in Box 1A. It is useful to distinguish between two types of mechanism for the formation of these structures [441]: self-assembly and self-organization. The former involves the aggregation of molecules such as proteins or lipids into a structure at thermodynamic equilibrium. On the other hand, the self-organization of nonequilibrium systems involves the spontaneous formation of a macroscopic structure based on the interactions between its microscopic components; there is no explicit architectural blueprint for the emerging structure [477]. Although many cellular structures appear static, a closer look reveals that they are highly dynamic systems, continuously exchanging proteins and other molecules with the surrounding environment. Self-organization can then occur when this is combined with nonlinear interactions between the various components of the given cellular structure.

Probably the most studied example of a self-organizing structure in cells is the *cytoskeleton* [275], which provides the structural integrity of the cell and plays a crucial role in determining cell shape and polarity. It also forms the so-called spindle apparatus during cell division or mitosis and drives cell motility. The cytoskeleton consists of a network of biopolymers including F-actin and microtubules, which undergo a continuous turnover of their subunits by net polymerization at one end (the plus end) and net depolymerization at the other end (the minus end). Growth and shrinkage of different components of the cytoskeleton are regulated by nonlinear interactions with various proteins, resulting in a self-organizing structure that can assume different forms from the same basic building blocks, depending on the particular functional role [285, 442]. The self-organizing properties of microtubule networks have been elegantly demonstrated in vitro by simply combining tubulin,

microtubule motors, and ATP [467]. Depending on the relative concentrations of motors and tubulin, structurally different patterns (i.e., random networks, vortices, or asters) are formed. The outcome of the polymerization process is solely determined by the concentration of reactants and the kinetics of their interactions. Moreover, there are sharp transitions between distinct assembly patterns, and a diverse range of initial conditions can result in the same assembly pattern. These are both characteristic signatures of self-organization.

Another likely candidate for cellular self-organization is the set of organelles forming the secretory pathway, namely, the *endoplasmic reticulum* (ER) and the *Golgi apparatus*. Newly translated proteins enter the secretory pathway, where they are sorted and transported to appropriate sites on the plasma membrane. One of the significant features of the secretory pathway is that there is a constant active exchange of molecules between the ER and Golgi apparatus, which have different lipid and protein compositions [392, 393]. Such an exchange is mediated by motor-driven *vesicular transport*. Vesicles bud from one compartment or organelle, carrying various lipids and proteins, and fuse with another compartment. Transport in the anterograde direction has to be counterbalanced by retrograde transport in order to maintain the size of the compartments and to reuse components of the transport machinery. Since bidirectional transport would be expected to equalize the composition of both compartments, there has been considerable interest in understanding the self-organizing mechanisms that allow such organelles to maintain their distinct identities while constantly exchanging material.

The above two examples involve active transport processes. A different mechanism for self-organization within cells (and between cells) is the interplay between passive diffusion and nonlinear chemical reactions, as modeled using *reaction–diffusion equations*. This builds upon the seminal work on morphogenesis by Turing [646], who established the principle that two nonlinearly interacting chemical species differing significantly in their rates of diffusion can amplify spatially periodic fluctuations in their concentrations, resulting in the formation of a stable periodic pattern. The so-called Turing mechanism has subsequently been applied to a wide range of problems in cell and systems biology, as extensively reviewed by Murray [461]. Given the fact that various cellular processes involve relatively few molecules, it is important to understand how to incorporate the effects of intrinsic fluctuations in the number of reacting molecules into a reaction–diffusion system. This then raises the challenging mathematical problem of analyzing and simulating stochastic PDEs.

Box 1A. A cellular parts list.

The interior of a cell is crowded with macromolecules and a vast array of organelles. An organelle of a *eukaryotic* cell (cell with a nucleus) is a compartmentalized structure separated from the cytoplasm by lipid membrane. Some of the major organelles are listed below (see Fig. 1.3 and [3]).

this figure will be printed in b/w

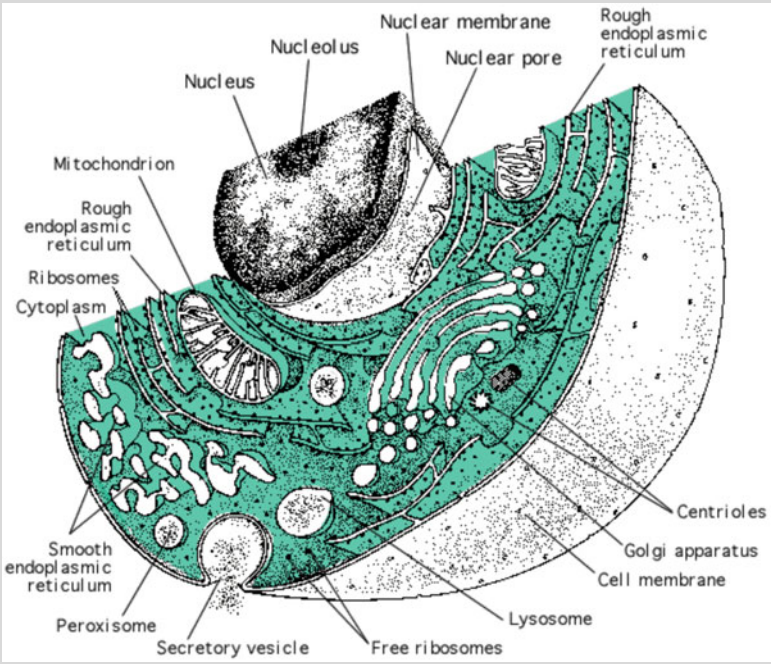


Fig. 1.3: The crowded environment of a eukaryotic cell (Public domain figure downloaded from Wikibooks)

1. *Nucleus.* The nucleus is a membrane-bound vesicle that contains the DNA genome. It is the site of transcription from DNA to mRNA via the action of polymerase; mRNA then exits the nucleus via nuclear pores and is then translated in the cytoplasm to produce proteins. Some of these proteins re-enter the nucleus as transcription factors (Sect. 6.1). In a human cell, the DNA molecules range in length from 1.9 to 8.5 cm, whereas the nucleus that contains 46 copies is an approximately spherical compartment with diameter usually less than 8 mm. Thus, each DNA copy must be reduced in length by more than a factor of 1,000. This is achieved using a hierarchical structure. First, the DNA is wrapped around *nucleosome* "core particles" resulting in the material *chromatin*. Fibers of the chromatin are then coiled and looped until each chromosome is only a few micrometers long and less than 1 mm thick. It follows that the nucleus itself is a highly crowded environment.

2. *Endoplasmic reticulum (ER).* The ER is one of the major organelles of the so-called secretory pathway. Proteins and lipids destined for the plasma membrane enter the ER from the nucleus as they are translated from mRNA by ER-associated ribosomes. The ER can be partitioned

into the rough ER (RER), which is rich in ribosomes, and the smooth ER (SER), which has only a few sparse ribosomes and tends to form a tubular structure. The ER is also an important intracellular store of Ca^{2+} (Sect. 3.4).

3. *Golgi apparatus.* An important aspect of the secretory pathway is that it is tightly regulated (Sect. 7.5). Proteins accumulate at specific exit sites and leave the ER in vesicles that transfer the cargo to organelles forming the Golgi network, where final packaging and sorting for target delivery is carried out. In most eukaryotic cells the Golgi network is confined to a region around the nucleus known as the Golgi apparatus, whereas in neurons there are Golgi "outposts" distributed throughout the dendrite.
4. *Mitochondria.* These are the major sites of ATP synthesis, which is the main source of biochemical energy at the macromolecular level.
5. *Lysosomes.* These play a major role in the degradation of cellular components and are filled with acids and degradative enzymes.

1.2 Experimental Studies of Noise in Cells

One of the major factors stimulating the recent interest in stochastic aspects of cellular function has been the rapid advance in experimental techniques for imaging and probing cells at the molecular level. Here we briefly review some of the methods most relevant to the biological themes covered in subsequent chapters. More details and examples can be found in Alberts et al. [3] and Phillips et al. [509].

1.2.1 FRAP and Single-Particle Tracking

A well-established experimental method for measuring protein mobility in cell membranes is *fluorescence recovery after photobleaching* (FRAP) [17, 532] (see Fig. 1.4). In this method, molecules of interest are tagged with a fluorophore such as a genetically encoded fluorescent protein. A small area of the cell is quickly photobleached using an intense beam of light, and the rate of fluorescence recovery in the excited spot is monitored over time. The recovery of the fluorescence signal depends on a variety of factors, including protein mobility and its availability from the area surrounding the bleached spot, the number of protein binding sites in the bleached area, and the rate of dissociation of bleached molecules from these binding sites. One of the limitations of FRAP is that its spatial resolution is limited by the diffraction of light to hundreds of nanometers. Moreover, it only gives the average

behavior of a population of molecules, so it cannot identify heterogeneous behavior at the level of individual molecules. Finally, mathematical modeling is often needed to help separate out the various factors contributing to recovery in order to measure important biophysical quantities such as the diffusivity.

In recent years a powerful alternative method has been developed based on *single-particle tracking* (SPT), in which one images the trajectory of a marker attached to a diffusing molecule [363, 566, 570]. Various transport properties of the particle are then derived through a statistical analysis of the trajectory, includ-

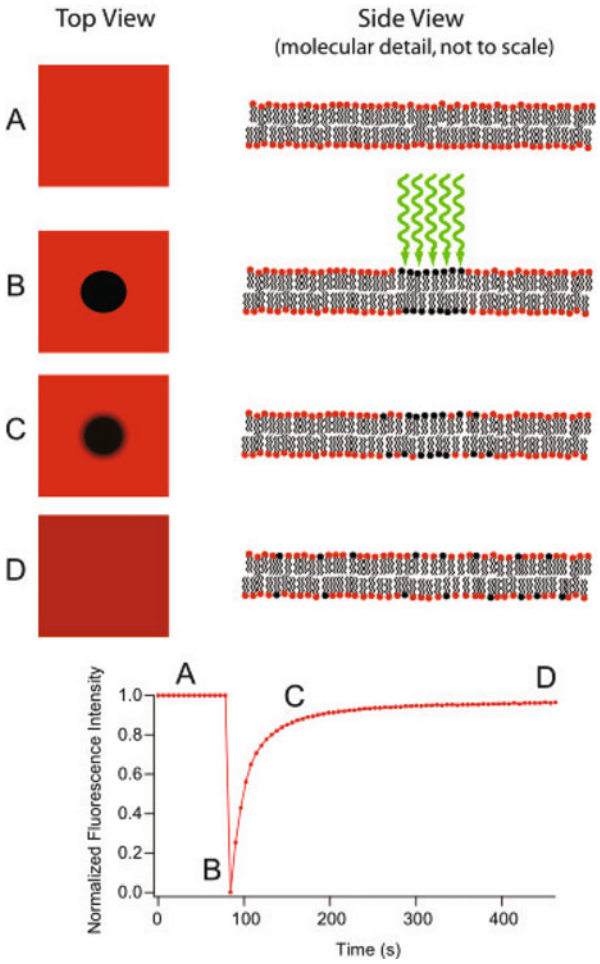


Fig. 1.4: Schematic illustration of fluorescence recovery after photobleaching (FRAP). (**a**) A membrane bilayer is uniformly labeled with a fluorescent tag. (**b**) This label is selectively photobleached by a small (30 µm) fast light pulse. (**c**) The intensity of the fluorescent signal within the bleached area is monitored as the bleached dye diffuses out and new dye diffuses in (**d**). Eventually uniform intensity is restored (Public domain figure downloaded from Wikipedia Commons)

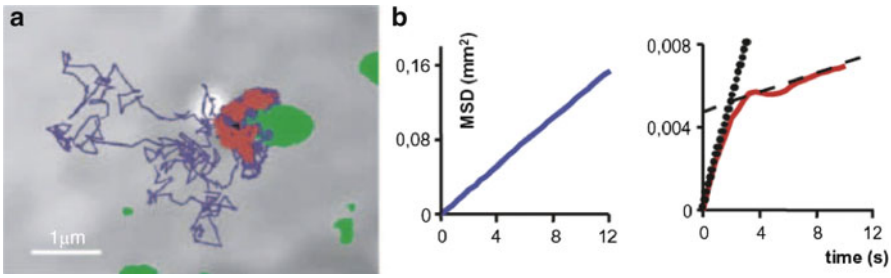


Fig. 1.5: Membrane receptor diffusion in neurons measured by single-particle tracking. (**a**) Superimposed image of the trajectory of 500 nm beads bound to glycine receptors (GlyRs) with the fluorescent image (*green*) of green fluorescent protein (GFP)-tagged gephyrin. Periods of free diffusion and confinement are indicated by *blue* and *red lines*, respectively. (**b**) Plots of the average mean-squared displacement (MSD) function during periods of free diffusion (*left panel*) and confinement (*right panel*) for GlyRs. Note the difference in both shape and amplitude of the MSDs. The *curved shape* of the MSD is characteristic of movement in a confined space (Adapted from [111])

ing a measurement of the mean-square displacement (MSD). Visualization of the diffusive behavior of single-membrane proteins in living cells has revealed that these molecules undergo a variety of stochastic behaviors including normal and anomalous diffusion and confinement within subcellular compartments. SPT also provides information on the structure of the surrounding membrane and the molecular interactions. The rapid increase in the range of applications of STP to cell biology has been driven by major improvements in the visualization of trajectories combined with new strategies for labeling proteins with nanoprobes. However, single-molecule approaches still have their own limitations, such as the shortness of observation times and the possibility that identified molecules are not representative of the population, which can lead to sampling errors. Hence, bulk methods such as FRAP and single-molecule methods are complementary experimental approaches to studying the molecular physiology of cells. Figure 1.5 illustrates one application of SPT, namely, studying the role of lateral membrane diffusion in delivering neurotransmitter receptors to synapses of a neuron [643].

1.2.2 Optical Tweezers or Traps

In recent years there have been spectacular advances in single-molecule techniques for measuring the force-dependent dynamics of molecular motors, DNA, and other macromolecules vital for cell function. In particular, the use of an *optical tweezer* (or trap) allows piconewton forces to be applied to molecules over nanometer length scales. The basic idea of an optical tweezer is to use radiation pressure from individual photons emitted by a laser to generate forces on a micron-sized glass

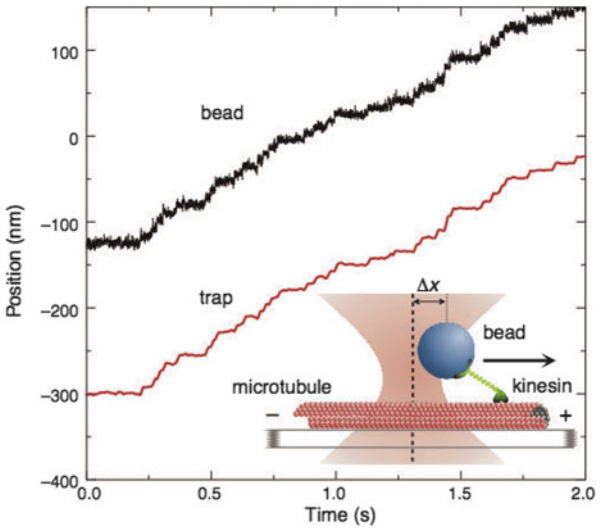


Fig. 1.6: Operation of an optical tweezer or trap. Experimental traces show kinesin-driven bead movement and the corresponding optical trap displacement; the separation between bead and trap was maintained at around 175 nm. The movement can be characterized in terms of a stochastic sequence of steps of average length 8 nm. *Inset*: schematic representation of the experimental setup (Adapted from [571])

bead. This then imparts a force on the macromolecule of interest via a velcro-like link between the molecule and the glass bead. By applying known forces to the bead, it is possible to study the mechanochemistry of the attached macromolecule as a function of the applied force. An example of such an experiment is illustrated in Fig. 1.6. Here a kinesin molecular motor attached to a silica bead moves along a clamped microtubule filament [571]. The bead is placed inside an optical trap such that the force on the bead is proportional to its displacement Δx from the center of the trap. The applied force is maintained at a constant level by using feedback to move the trap in sync with the bead so that Δx is kept constant. Such an experiment can be used to determine the probability distribution of motor step lengths, for example. Another common application of the optical tweezer involves attaching the glass bead to one end of a DNA strand. This can then be used to determine the force-extension curve of a DNA polymer, measure the force produced by RNA polymerase (RNAP) during the transcription of DNA, or measure the force necessary to pack viral DNA into the capsid (protein shell) of a bacteriophage (virus that infects and replicates within bacteria).

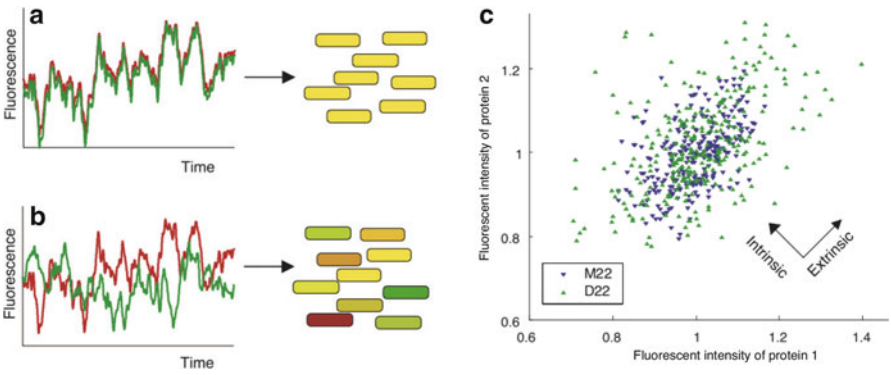


Fig. 1.7: Measuring intrinsic and extrinsic noise in gene expression. Two almost identical genes, which encode *red* and *green* fluorescent proteins, are expressed from identical promoters and are influenced identically by cell-specific factors, such as gene-regulatory signals. (**a**) Cells with equal amounts of the two proteins appear *yellow*, indicating that the level of intrinsic noise is low. Noise fluctuations of the two proteins in the same cell appear correlated over time. (**b**) If intrinsic noise is significant, then the expression of the two genes becomes uncorrelated in individual cells, giving rise to a cell population in which some cells express more of one fluorescent protein than the other. (**c**) Plot of fluorescence in two strains (M22 and D22) of the bacterium *Escherichia coli*. Each point represents the mean fluorescence intensities from one cell. Spread of points perpendicular to the diagonal line on which the two fluorescent intensities are equal corresponds to intrinsic noise, whereas the spread parallel to this line corresponds to extrinsic noise (Adapted from [164])

1.2.3 Two-Reporter Assays

As was mentioned in Sect. 1.1, there is an important distinction between intrinsic and extrinsic noise. Elowitz et al. [164] developed a *two-reporter assay* that can discriminate between the two within the context of gene expression. (A biochemical assay is an experimental procedure for quantitatively measuring the presence or amount of one or more target molecular constituents.) In this particular assay, two almost identical fluorescent proteins are simultaneously expressed from two genes that are controlled by identical regulatory sequences (the same promoter). Cells with the same amount of each protein appear yellow, whereas cells expressing more of one fluorescent protein than the other appear green or red (see Fig. 1.7a,b). In the absence of intrinsic noise, the expression of the two-reporter proteins should be strongly correlated. On the other hand, since the expression of the two reporters is independent, any intrinsic stochasticity in gene expression will be manifested as differences in expression levels within the same cell. By considering the spread of the expression levels across a population of cells, it is possible to separate out the noise contribution generated by the biochemical reaction steps that are intrinsic to the process of gene expression from extrinsic environmental noise (Fig. 1.7c).

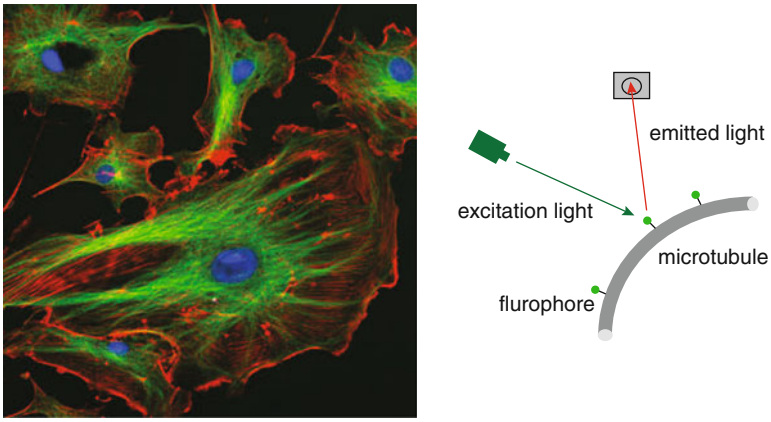


Fig. 1.8: The cytoskeleton of a eukaryotic cell revealed by fluorescent imaging. Actin filaments are shown in *red*, microtubules in *green*, and the nuclei are in *blue* (Public domain figure downloaded from Wikipedia Commons)

1.2.4 Fluorescent Imaging

In *fluorescent imaging* a target structure such as a microtubule filament is labeled with fluorophores. Incident photons of a given wavelength are absorbed by the fluorophores. The latter become energetically excited and emit light at a different wavelength, which is then detected. The illumination light is separated from the much weaker emitted fluorescence through the use of a spectral emission filter. The use of selective labeling means that only the structures of interest are observed under the microscope. Hence, one can measure the size and number of various structures based on the strength of the fluorescent signal (see Fig. 1.8). One can also use time-lapse microscopy to track changes in the shape or size of a cell, which is crucial for understanding processes such as cell motility and cell polarization. Another important application of time-lapse fluorescent microscopy is to obtain statistical data regarding the active intracellular transport of vesicles. By tagging proteins or mRNA contained within the vesicle it is possible to monitor the bidirectional transport of vesicles along the axons and dendrites of neurons (for example, see Fig. 1.9).

1.3 A Brief Introduction to Probability Theory

Since this is a book on stochastic processes, we thought that it would be helpful to summarize some basic concepts in probability theory. A more abstract formulation can be found in Chap. 11.

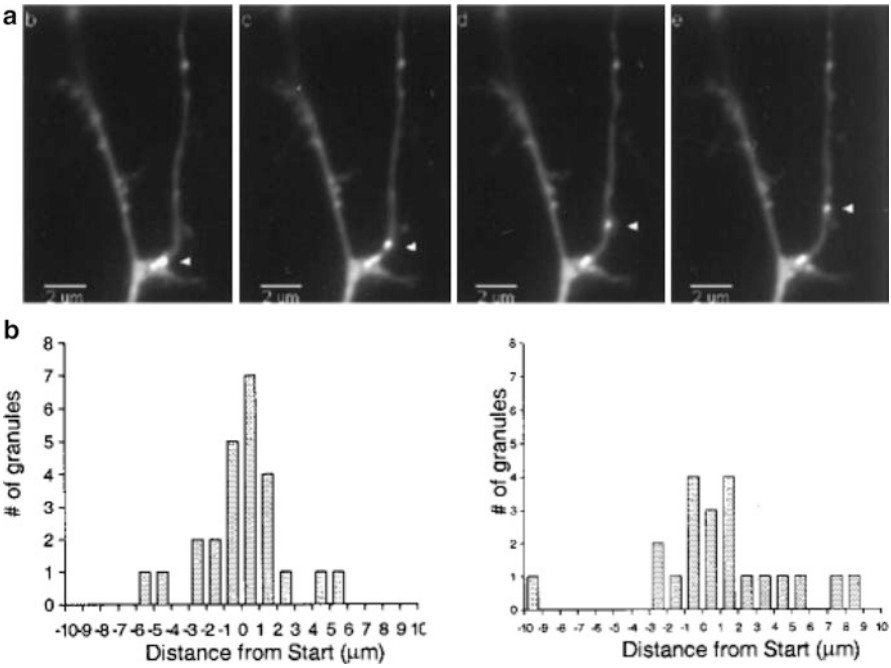


Fig. 1.9: (**a**) Time-lapse fluorescent images of an actively transported vesicle (*arrowhead*) within the dendrite of a hippocampal neuron. The vesicle contains fluorescently labeled mRNA for the kinase CaMKII. Frames are sequential images taken every 20 s. The total distance covered was 5.85 mm and the average velocity over 160 s was 0.04–0.01 mm/s. (**b**) The distance traveled by vesicles is given by a Gaussian-like histogram. If the neuron is depolarized by treating with KCl, then the histogram is shifted towards anterograde (away from the cell body) movement (Adapted from [544])

1.3.1 Random Variables

Let us start with the idea of a random variable X, which is an unknown quantity that has one of a number of possible values with some relative likelihood. For example, in the case of an unbiased die, the possible values of X are the integers 1–6 and the probability of throwing any one of these is $p = 1/6$. From a frequentist interpretation of probability, this means that in the limit that the number of throws $N \to \infty$, the fraction that will have the value four, say, is equal to $1/6$. More, generally, suppose that $X = n$ with probability p_n for $n = 1, \ldots, K$. Since X takes on discrete values it is called a *discrete random variable*. Given that X must take on one of these values, the total probability must be 1, that is,

$$\sum_{n=1}^{K} p_n = 1.$$

Two important statistical quantities are the *average, mean, or expectation value* of X, which is defined by

$$\mathbb{E}[X] \equiv \langle X \rangle = \sum_{n=1}^{K} n p_n,$$

and the *variance* of X,

$$\mathrm{var}[X] \equiv \langle X^2 \rangle - \langle X \rangle^2 = \sum_{n=1}^{K} n^2 p_n - (\sum_{n=1}^{K} n p_n)^2.$$

The *standard deviation* of X, denoted by σ_X, is defined to be $\sigma_X = \sqrt{\mathrm{var}[X]}$ and is a measure of how broad the probability distribution of X is. These ideas carry over to the case of a *continuous random variable* X that can take any value on the real line $\mathbb{R}$. The relative likelihood of the different values of X is now given by a *probability density* $p(x)$ such that

$$\mathrm{Prob}[a < x < b] = \int_a^b p(x) dx.$$

The total probability is again equal to 1, which can be expressed as the normalization condition

$$\int_{-\infty}^{\infty} p(x) dx = 1.$$

Similarly, the mean and variance are defined according to

$$\langle X \rangle = \int_{-\infty}^{\infty} x p(x) dx,$$

and

$$\mathrm{var}[X] = \int_{-\infty}^{\infty} x^2 p(x) dx - \langle X \rangle^2.$$

In the following, we will focus on continuous random variables, although analogous results hold for discrete random variables.

1.3.2 Conditional Probabilities and Correlations

Two random variables X and Y are said to be *independent* if the relative likelihood of a particular value taken by one variable is unaffected by the particular value taken by the other variable. This means that the joint probability density $p(x,y)$ can be written as the product of individual probability densities for each of the random variables, $p(x,y) = p_X(x) p_Y(y)$, and

$$\begin{aligned}\text{Prob}[a < X < b \text{ and } c < Y < d] &= \int_a^b \int_c^d p(x,y)dydx \\ &= \int_a^b \int_c^d p_X(x)p_Y(y)dydx = \left(\int_a^b p_X(x)dx\right)\left(\int_c^d p_Y(y)dy\right) \\ &= \text{Prob}[a < X < b] \cdot \text{Prob}[c < Y < d].\end{aligned}$$

It immediately follows that $\langle XY\rangle = \langle X\rangle\langle Y\rangle$. If, on the other hand, X and Y are dependent, then $p(x,y)$ cannot be decomposed into such a product. Nevertheless, one can define a so-called *marginal* probability density for X, say, which applies when there is no information about the value of Y:

$$p_X(x) = \int_{-\infty}^{\infty} p(x,y)dy.$$

Dependence suggests that if the value of Y is known, then the probability density for X will be modified. The probability density for $X = x$ *given* that $Y = y$ is denoted by the *conditional* probability density $p(x|y)$. A basic result of conditional probability densities is

$$p(x,y) = p(x|y)p_Y(y) = p(y|x)p_X(x).$$

It follows that

$$p(y|x) = \frac{p(x|y)p_Y(y)}{\int_{-\infty}^{\infty} p(x|y)p_Y(y)dy},$$

which is a statement of *Bayes' theorem* [402]. Although we will mainly use the frequency interpretation of probability in this book, we note that the Bayesian interpretation is often used in statistical and information theoretic approaches to systems biology (Sect. 6.5). For example, suppose that X represents some observable data and Y represents a parameter of the system that produces the data. Then $p(x|y)$ is known as the likelihood function that a particular parameter value produces the observed data and $p_Y(y)$ is the Bayesian prior, which expresses what was known (or thought to be known) about the parameter before the measurement. Finally, $p(y|x)$ is the updated posterior probability density for the parameter obtained by combining the prior information with the information gained from the measurement.

An important quantity that indicates the degree of mutual dependence of two random variables is the *correlation coefficient* of X and Y, defined by

$$C_{XY} = \frac{\langle XY\rangle - \langle X\rangle\langle Y\rangle}{\sqrt{\text{var}[X]\text{var}[Y]}}.$$

The numerator is known as the *covariance* of X and Y and vanishes if X and Y are independent. On the other hand, if X and Y are perfectly correlated ($X = cY, c > 0$), then $C_{XY} = 1$, and if they are perfectly anticorrelated ($X = cY, c < 0$), then $C_{XY} = -1$.

1.3.3 Adding Random Variables

Suppose that N independent measurements of some quantity are made, which are denoted by X_n, $n = 1,\ldots,N$. The random variables are said to be *independent, identically distributed* (i.i.d.) random variables and have the same mean μ and variance σ^2. Averaging the results of these measurements generates a new random variable

$$X = \sum_{n=1}^{N} \frac{X_n}{N}.$$

It follows that

$$\langle X \rangle = \left\langle \sum_{n=1}^{N} \frac{X_n}{N} \right\rangle = \frac{1}{N} \sum_{n=1}^{N} \langle X_n \rangle = \mu$$

and

$$\begin{aligned}
\mathrm{var}[X] &= \left\langle \sum_{m=1}^{N} \frac{X_m}{N} \sum_{n=1}^{N} \frac{X_n}{N} \right\rangle - \mu^2 = \frac{1}{N^2} \sum_{m=1}^{N} \sum_{n=1}^{N} \langle X_m X_n \rangle - \mu^2 \\
&= \frac{1}{N^2} \sum_{m=1}^{N} \left[\langle X_m^2 \rangle + \sum_{n,n\neq m} \langle X_m X_n \rangle - N\mu^2 \right] \\
&= \frac{1}{N^2} \sum_{m=1}^{N} \left[\sigma^2 + \sum_{n,n\neq m} \langle X_m \rangle \langle X_n \rangle - (N-1)\mu^2 \right] = \frac{\sigma^2}{N}.
\end{aligned}$$

We thus obtain the well-known result that the standard deviation of the average varies as $\sigma_X \sim N^{-1/2}$. This is an expression of the *law of large numbers*, which states that the average of the results obtained from a large number of trials should be close to the expected value and will tend to become closer as more trials are performed.

A related result is the *central limit theorem* [242]. Suppose that $\{X_1, X_2, \ldots\}$ is a sequence of i.i.d. random variables with $\langle X_j \rangle = \mu$ and $\mathrm{var}[X_j] = \sigma^2 < \infty$. Then as n approaches infinity, the random variables $\sqrt{n}(S_n - \mu)$ with $S_n = \sum_{j=1}^{n} X_n/n$ converge in distribution to a normal or Gaussian distribution $N(0,\sigma^2)$:

$$\sqrt{n}\left(\frac{1}{n} \sum_{j=1}^{n} X_j - \mu \right) \xrightarrow{D} N(0,\sigma^2).$$

The Gaussian distribution $N(\mu,\sigma^2)$ is defined according to

$$N(\mu,\sigma^2)(x) = p(x) \equiv \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right).$$

Convergence in distribution means that

$$\lim_{n\to\infty} \text{Prob}[\sqrt{n}(S_n - \mu) \leq z] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{z} \mathrm{e}^{-x^2/2\sigma^2} dx.$$

1.3.4 Transformation of Random Variables

Suppose that we know the probability density $p_X(x)$ of a random variable X and we construct a new random variable $Y = g(X)$, where g is an invertible function. We would like to determine the probability density $P_Y(y)$. This can be achieved by considering the expectation value of a function $f(Y)$ in terms of $p_X(x)$ and performing a change of variables. That is,

$$\begin{aligned}\langle f(Y)\rangle &= \int_{x=a}^{x=b} f(g(x))p_X(x)dx = \int_{y=g(a)}^{y=g(b)} f(y)p_X(g^{-1}(y))\frac{dx}{dy}dy \\ &= \int_{y=g(a)}^{y=g(b)} f(y)\frac{p_X(g^{-1}(y))}{g'(x)}dy = \int_{y=g(a)}^{y=g(b)} f(y)\frac{p_X(g^{-1}(y))}{g'(g^{-1}(y))}dy.\end{aligned}$$

It is possible that on transforming the limits of the integral, $g(b) < g(a)$, which means that the fraction is then negative. Therefore, the transformed probability density is

$$p_Y(y) = \frac{p_X(g^{-1}(y))}{|g'(g^{-1}(y))|}.$$

1.3.5 Moments and Cumulants

Given a continuous random variable X with probability density $p(x)$, the expectation value of X^n, $\langle X^n\rangle$, is called the nth *moment*. The moments can be calculated using the so-called *characteristic function* of X, which is defined by

$$G(k) = \int_{-\infty}^{\infty} \mathrm{e}^{ikx} p(x)dx.$$

It can be seen that $G(k)$ is the Fourier transform of $p(x)$. Taylor expanding $G(k)$, we have

$$G(k) = \sum_{n=0}^{\infty} \frac{G^{(n)}(0)k^n}{n!},$$

where $G^{(n)}(k)$ is the nth derivative of $G(k)$. An alternative series expansion of $G(k)$ is obtained by noting that

$$G(k) = \left\langle \mathrm{e}^{ikx} \right\rangle = \left\langle \sum_{n=0}^{\infty} \frac{(ikX)^n}{n!} \right\rangle = \sum_{n=1}^{\infty} \frac{i^n \langle X^n \rangle k^n}{n!}.$$

Equating the two series representations of $G(k)$ shows that

$$\langle X \rangle^n = (-i)^n G^{(n)}(0).$$

A related quantity is the nth order *cumulant* of X, which we denote by κ_n. It is a polynomial in the first n moments, which is defined by

$$\kappa_n = (-i)^n \left. \frac{d^n}{dk^n} \ln G(k) \right|_{k=0}.$$

The first three cumulants are $\kappa_1 = \langle X \rangle$, $\kappa_2 = \mathrm{var}[X]$, and

$$\kappa_3 = \langle X^3 \rangle - 3\langle X^2 \rangle \langle X \rangle + 2\langle X \rangle^3.$$

One useful feature of cumulants is that $\kappa_n = 0$ for all $n \geq 3$ in the case of the Gaussian distribution. This implies that all higher moments of a Gaussian can be written in terms of the mean $\langle X \rangle$ and variance σ^2. A general formula for these moments can be derived using standard results from Gaussian integration, and one finds that

$$\langle (X - \langle X \rangle)^{2n} \rangle = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} y^{2n} \mathrm{e}^{-y^2/2\sigma^2} dy = \frac{(2n-1)!\sigma^{2n}}{2^{n-1}(n-1)!},$$

and

$$\langle (X - \langle X \rangle)^{2n-1} \rangle = 0$$

for $n \geq 1$. It also turns out that the characteristic function of a Gaussian is also a Gaussian, since

$$G(k) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \mathrm{e}^{ikx} \mathrm{e}^{-x^2/2\sigma^2} dx = \frac{e^{-\sigma^2 k^2/2}}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \mathrm{e}^{-(x - i\sigma^2 k)^2/2\sigma^2} dx = e^{-\sigma^2 k^2/2}.$$

(Technically speaking, the Gaussian integral is evaluated by completing the square and then using analytical continuation in the complex k-plane.)

1.3.6 Stochastic Processes

One final point to make is that this book is concerned with stochastic processes, which involve random variables evolving in time. Thus a random variable will have an additional time label $X \to X_n, n \in \mathbf{Z}^+$ for discrete-time processes and $X \to X(t), t \in \mathbb{R}^+$ for continuous-time processes. Roughly speaking, one can treat t (or n) as a parameter so that for fixed t, $X(t)$ is a random variable in the sense considered in this section. However, various objects such as the probability density and characteristic function are now parameterized by t and we write $p(x) \to p(x,t)$ and $G(k) \to G(k,t)$ etc.

1.4 Statistical Mechanics and the Boltzmann–Gibbs Distribution

Here we collect together some fundamental results in statistical mechanics that commonly arise when considering stochastic processes within cells. We will introduce quantities such as free energy, entropy, and chemical potential from the perspective of the Boltzmann–Gibbs distribution of a macromolecule such as DNA or a protein in thermodynamic equilibrium at temperature T, as illustrated in Fig. 1.2. For a much more general and detailed treatment of statistical mechanics from the physics perspective, see [102, 317]. A more biophysical approach can be found in [295, 509]. Note that the only topics in the book where statistical mechanics is needed are the equilibrium theory of polymers (Sect. 4.5), the fluctuation–dissipation theorem (Sect. 5.2), receptor clustering (Sect. 5.3), information transmission in gene networks (Sect. 6.5), and the translocation of polymers through membranes (Sect. 7.3.4). Otherwise, one only needs to be familiar with the definition of the Boltzmann–Gibbs distribution in Eqs. (1.4.1) and (1.4.5), so that the remainder of this section could be skipped.

1.4.1 Boltzmann–Gibbs Distribution

Suppose that a macromolecule such as a protein or DNA has a set of internal microstates labeled by j and let E_j denote the energy of the molecule in the jth state. (These could include different folded or twisted configurations of the underlying amino acid or nucleotide chain.) As we briefly mentioned in Sect. 1.1, a fundamental principle of equilibrium statistical mechanics is that the probability p_j that the molecule is in state j is given by the *Boltzmann–Gibbs distribution*

$$p_j = \frac{1}{Z} \mathrm{e}^{-E_j/k_B T}, \quad Z = \sum_j \mathrm{e}^{-E_j/k_B T}, \tag{1.4.1}$$

where T is the temperature of the surrounding cellular environment and Z is the partition function. As a simple example, consider a molecule that can exist in two states $j = C, O$ with energies E_C and E_O, respectively. The associated probabilities are

$$P_O = \frac{\mathrm{e}^{-E_O/k_B T}}{e^{-E_O/k_B T} + e^{-E_C/k_B T}} = \frac{1}{1 + \mathrm{e}^{\Delta E/k_B T}}, \quad P_C = 1 - P_O,$$

with $\Delta E = E_O - E_C$ the difference in energies between the two states. It follows that

$$\frac{P_O}{P_C} = \mathrm{e}^{-\Delta E/k_B T}.$$

Suppose that the system approaches thermodynamic equilibrium according to the two-state Markov process

$$\frac{dP_C}{dt} = -\alpha P_C + \beta P_O \quad \frac{dP_O}{dt} = \alpha P_C - \beta P_O. \tag{1.4.2}$$

This has the unique stable steady state

$$P_O = \frac{\alpha}{\alpha+\beta}, \quad P_C = \frac{\beta}{\alpha+\beta}. \tag{1.4.3}$$

Such a steady state has to be consistent with equilibrium statistical mechanics, which implies that the forward and backward transition rates have to satisfy the condition

$$\frac{\alpha}{\beta} = \mathrm{e}^{-\Delta E/k_B T}. \tag{1.4.4}$$

Examples of a two-state system include an ion channel that is either open or closed and a protein receptor that is either bound or unbound to a ligand (see Sect. 3.1).

1.4.2 Free Energy and Entropy

The Boltzmann–Gibbs distribution also applies to continuous states of a molecule. For example, consider a molecule moving in $d = 1,2,3$ dimensions under the influence of a conservative force—a force that can be written as $\mathbf{f} = -\nabla\Phi(\mathbf{x})$ where $\Phi(\mathbf{x})$ is some potential energy function. Examples include gravitational and electrical potentials. Assuming that all other degrees of freedom are independent of $\mathbf{x}$ (momentum, internal microstates), the equilibrium probability density with respect to position $\mathbf{x}$ is given by

$$p(\mathbf{x}) = \frac{1}{Z}\mathrm{e}^{-\Phi(\mathbf{x})/k_B T}, \quad Z = \int \mathrm{e}^{-\Phi(\mathbf{x})/k_B T} d\mathbf{x}. \tag{1.4.5}$$

The situation becomes more interesting when one cannot assume separability between different degrees of freedom. In particular, suppose that the total energy of a molecule is independent of its internal state j, but the number of internal states $\Omega(\mathbf{x})$ depends on position. If $p_j(\mathbf{x})$ is the joint probability density of being at position $\mathbf{x}$ and in internal state j, then the marginal probability density $p(\mathbf{x})$ is given by

$$p(\mathbf{x}) = \sum_j p_j(\mathbf{x}) = \frac{1}{Z}\sum_j \mathrm{e}^{-\Phi(\mathbf{x})/k_B T} = \frac{1}{Z}\Omega(\mathbf{x})\mathrm{e}^{-\Phi(\mathbf{x})/k_B T},$$

with

$$Z = \int \sum_j e^{-\Phi(\mathbf{x})/k_B T} d\mathbf{x} = \int \Omega(\mathbf{x})\mathrm{e}^{-\Phi(\mathbf{x})/k_B T} d\mathbf{x}.$$

If we now introduce the quantity

$$E(\mathbf{x}) = \Phi(\mathbf{x}) - k_B T \ln \Omega(\mathbf{x}), \tag{1.4.6}$$

then we can express $p(\mathbf{x})$ in terms of the effective Boltzmann–Gibbs distribution

$$p(\mathbf{x}) = \frac{1}{Z} \mathrm{e}^{-E(\mathbf{x})/k_B T}, \quad Z = \int \mathrm{e}^{-E(\mathbf{x})/k_B T} d\mathbf{x}.$$

Another major result from statistical mechanics is that one should treat $E(\mathbf{x})$ as the effective energy of the molecule in thermodynamic equilibrium. One defines $E(\mathbf{x})$ to be the *free energy* of the molecule and the term $S = k_B \ln \Omega(\mathbf{x})$ to be the *entropy*. An immediate consequence of the above result is that the total force on the molecule is

$$\mathbf{f}_{\text{tot}}(\mathbf{x}) = -\nabla E(\mathbf{x}) = -\nabla \Phi(\mathbf{x}) + T \nabla S(\mathbf{x}) = f(\mathbf{x}) + T \nabla S(\mathbf{x}). \tag{1.4.7}$$

That is, there is an additional *entropic force* given by $T\nabla S(\mathbf{x})$. The entropic force has the following statistical mechanical interpretation: the total entropy of a closed system cannot decrease (second law of thermodynamics), so that if a change in position $\delta\mathbf{x}$ decreases the entropy of the molecule, $\nabla S(\mathbf{x}) \cdot \delta\mathbf{x} < 0$, then this results in the environment having to do work $\delta W = -T \nabla S(\mathbf{x}) \cdot \delta\mathbf{x} > 0$ to counteract the entropic force. Since the internal energy of the molecule does not change when it is displaced, the work done is dissipated as heat (conservation of energy), resulting in an increase in the environmental entropy, thus counterbalancing the decrease in entropy of the molecule. (For simplicity, we are treating the molecule and its surrounding environment as a closed system.) One important example of an entropic force arises in the uncoiling of a flexible polymer (see Sect. 4.5).

1.4.3 Chemical Potential of a Solution

Another important consequence of entropic effects arises from changes in the number of microstates when a solute molecule is removed from a dilute solution. Suppose that there are N solvent + solute molecules and n solute molecules with $n \ll N$. For simplicity, let us represent the solution in terms of N boxes that can either be occupied by a solute molecule or a solvent molecule (see Fig. 1.10). The number of different configurations for given n, N is given by the combinatorial factor for distributing n items in N boxes:

$$\Omega(n) = \frac{N!}{n!(N-n)!}.$$

Taking logs and using Stirling's formula

$$\log N! \approx N\log N - N + \frac{1}{2}\log(2\pi N),$$

we have the entropy

$$S(n) = k_B[N\ln N - n\ln n - (N-n)\ln(N-n)].$$

Thus there is an entropic contribution to the free energy of the solute of the form $-k_BT\ln S(n)$. Now suppose that one solute molecule is removed from solution by binding to a protein receptor embedded in the cell membrane, for example. The change in free energy consists of two contributions: the change in energy ε associated with binding to the receptor and the change in entropy of the solution due to $n \to n-1$. Thus the total change in free energy (for $N \gg n$) is

$$\Delta E = \varepsilon + T\frac{dS(n)}{dn} \approx \varepsilon - k_BT\ln(n/N).$$

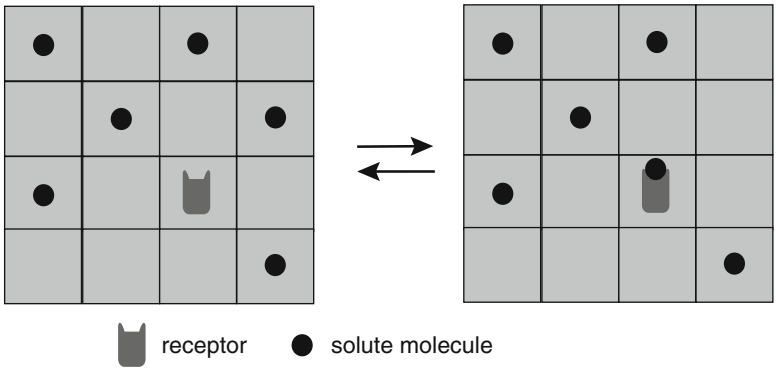


Fig. 1.10: Lattice model of a receptor binding with a solute molecule

Typically, one transforms n/N to a volume concentration c and sets

$$\Delta E = \varepsilon - \mu, \quad \mu = \mu_0 + k_BT\ln(c/c_0), \tag{1.4.8}$$

where c_0 is a reference concentration and μ_0 is a constant. The quantity μ is called the *chemical potential* of the solute. One application of the chemical potential is to ligand-gated ion channels. Suppose that the ion channel receptor can be in two

states: unbound by a ligand (C) and bound by a ligand (O). Using the Boltzmann–Gibbs distribution for a two-state ion channel with $E_C = \mu_0, E_O = \Delta E$ implies that

$$P_O = \frac{\mathrm{e}^{-(\varepsilon-\mu+\mu_0)/k_BT}}{1+\mathrm{e}^{-(\varepsilon-\mu+\mu_0)/k_BT}} = \frac{(c/c_0)\mathrm{e}^{-\varepsilon/k_BT}}{1+(c/c_0)\mathrm{e}^{-\varepsilon/k_BT}}. \tag{1.4.9}$$

This result can be used to determine the so-called dissociation constant that appears in chemical mass-action kinetics (Sect. 3.1).

1.5 Organization of the Book

We now give a brief overview of the organizational structure of the book. It is divided into two parts, with Chaps. 2–6 covering foundations and Chaps. 7–11 covering more advanced topics. Chapter 2 develops two microscopic theories of diffusion in cells, one based on random walks and the other on overdamped Brownian motion. The latter introduces the theory of continuous stochastic processes, SDEs (Langevin equations), and the FPE. The important topics of first passage times (FPTs) and diffusion-limited reactions are also described. Chapter 3 considers the problem of stochastic ion channel gating, introducing the theory of biochemical reaction kinetics and the law of mass action, discrete Markov processes, and master equations. It is also shown how a discrete process can be approximated by a continuous process using a system-size expansion of the master equation. The resulting FPE is then used to study some escape problems in bistable stochastic ion channel models, including calcium-induced calcium release (CICR) and spontaneous action potential (SAP) generation in neurons. Chapter 4 describes how random walks and SDEs are used to model polymerization and molecular motors. Polymerization plays a major role in the self-organization of cytoskeletal structures, whereas molecular motors "walking" along polymer filaments are a major active component of intracellular transport. Chapters 5 and 6 consider stochastic biochemical reaction networks. Chapter 5 focuses on how cells sense biochemical cues in a noisy environment, addressing the fundamental physical limits of biochemical signaling and mechanisms for amplifying and adapting signals. Applications to bacterial chemotaxis and active mechanotransduction in hair cells are considered. Chapter 6 covers a wide range of topics in stochastic gene expression, including transcriptional and translational bursting, autoregulatory networks and transcription factors, genetic switches and the *lac* operon, genetic oscillators and circadian rhythms, information transmission by gene networks, and kinetic proofreading. Part I also provides background material in applied mathematics that is particularly relevant to the themes of this book: Fourier and Laplace transforms, the Dirac delta function, linear stability analysis, Hopf bifurcations, and methods for solving linear PDEs such as separation of variables, eigenfunction expansions, and Green's functions.

Part II begins with Chap. 7 on passive (diffusion-driven) and active (motor-driven) mechanisms for intracellular transport. Topics include anomalous diffusion in the crowded cellular environment, diffusion in membrane channels and pores, PDE models of active transport, molecular exclusion processes, and random intermittent search processes. Chapters 8 and 9 explore various examples of self-organization in cells. First, the roles of polymerization in cellular length regulation, cell mitosis, and cell motility are considered, and then reaction–diffusion models are used to study intracellular protein concentration gradients, intracellular pattern formation, and cell polarization. Finally, Chaps. 10 and 11 consider more advanced topics in probability theory and stochastic processes relevant to cell biology. First, it is shown how Wentzel–Kramers–Brillouin (WKB) methods and large deviation theory can be used to obtain more accurate solutions of noise-induced escape problems in bistable systems, compared to the diffusion approximation used in Part I. Second, a more abstract formulation of probability theory is introduced, in particular, the theory of discrete martingales, which plays an important role in the development of numerical methods for simulating chemical reaction networks. At the end of each chapter there is a set of exercises that further develops the mathematical models and analysis introduced within the body of the text. Additional background material is contained in gray boxes scattered throughout the text.

1.5.1 How to Use the Book

Part I can be used to teach a one semester advanced undergraduate or graduate course on stochastic processes in cell biology. From a biological perspective, these chapters cover the basic molecular components of the cell: ion channels, receptors, biopolymers, molecular motors, biochemical signaling networks, and gene networks (see Fig. 1.11). From a mathematical perspective, Part I develops the basic techniques needed to analyze stochastic processes in cell biology. Here we briefly outline these methods (see Fig. 1.12).

Continuous Markov Processes. There are two complementary approaches to formulating continuous stochastic process, one in terms of the sample paths generated by an SDE or Langevin equation and the other in terms of the FP equation describing the evolution of the probability density of possible paths. The former requires at least a basic understanding of stochastic calculus, namely, the rules for integrating an SDE in order to obtain an expression that can be used to generate moments of the stochastic process (Sects. 2.2 and 2.6). In the case of linear SDEs one can also use Fourier methods to determine the power spectrum, for example, which is important in quantifying the linear response properties of a noisy system (Sect. 2.2.5). The FP equation is a deterministic PDE that generalizes the diffusion equation and can be analyzed using standard methods in the theory of linear PDEs: separation of variables, boundary value problems, transform methods, Green's functions, and eigenfunction expansions (Sect. 2.5). A less standard technique is needed to deal with an FP equation for a Brownian particle moving in a periodic potential, which

is important in the theory of molecular motors (Sects. 4.2 and 4.3). (It also arises in studies of diffusion in narrow channels; see Sect. 7.3.)

Discrete Markov Processes. The probability distribution of a discrete Markov process evolves according to a system of coupled ODEs known as a master equation. In certain cases, the steady-state distribution can be solved by iterating a difference equation, and the full time-dependent master equation can be solved using discrete Fourier and Laplace transforms (Sects. 2.1, 3.2, 4.1, 6.2, 6.3). The resulting characteristic function or generating function can be used to generate moments of the distribution, in which case it is not necessary to determine the inverse transform. However, in general it is not possible to obtain an exact solution of a master equation, and so some form of approximation scheme is needed. The most common method is the so-called system-size expansion, which reduces the master equation to an FP equation (Sects. 3.2, 6.3, 6.4). One possible limitation of the reduction is that it assumes that the number of molecules involved is sufficiently large. Since many biochemical networks involve small numbers of molecules, this can result in significant errors. (A more accurate approximation scheme is to use WKB methods (Chap. 10).)

FPTs and Escape Problems. Many quantities measured by experimentalists can be interpreted mathematically in terms of the solution to a FPT problem—switching times of bistable gene networks, dwell times of a molecular motor stepping along a filament, release times of calcium sparks, firing times of neurons, etc. In the case of a continuous Markov process, the distribution of FPTs satisfies a differential equation that can be derived from the FP equation and then solved to determine the mean FPT, for example; such methods can also be applied to discrete Markov processes using a system-size expansion (Sects. 2.3, 3.3–3.5).

Statistical Mechanics and Information Theory. It is often possible to treat a molecule within a cell as in local thermodynamic equilibrium so that its properties can be characterized in terms of the Boltzmann–Gibbs distribution. How it interacts with the surrounding environment then requires a basic understanding of free energy and entropy. The ideas briefly introduced in Sect. 1.4 are further developed within the context of polymers in Sect. 2.5 and applied to the analysis of receptor clustering in Sect. 5.3. Another important application of statistical mechanical methods is to information theory, which is concerned with measuring how effective some biochemical or gene network is in transmitting information from the environment in the presence of noise (Sect. 6.5).

Part II can be used as a supplement to Part I, either as a reference or as a source of projects that follow up the ideas and methods introduced in Part I. Chapters 7–9 focus on more spatial aspects of cell biology including transport processes and self-organization: anomalous diffusion, membrane transport, cell mitosis, cell motility, and cell polarization. Chapters 10 and 11 develop more advanced methods in stochastic processes: WKB approximation, large deviations, path integrals, probability spaces, and martingales.

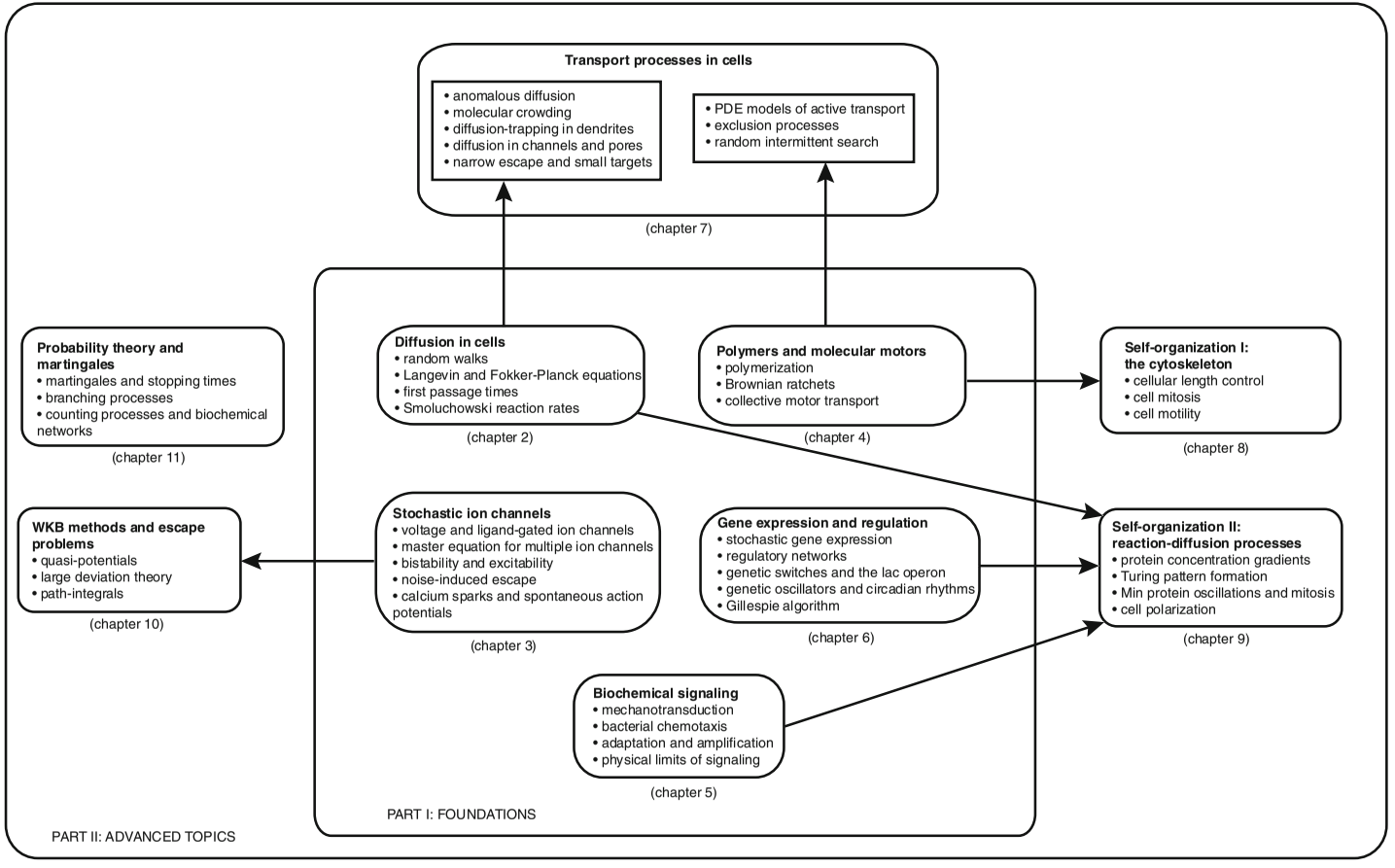


Fig. 1.11: Organizational flow chart of the book contents based primarily on biological themes

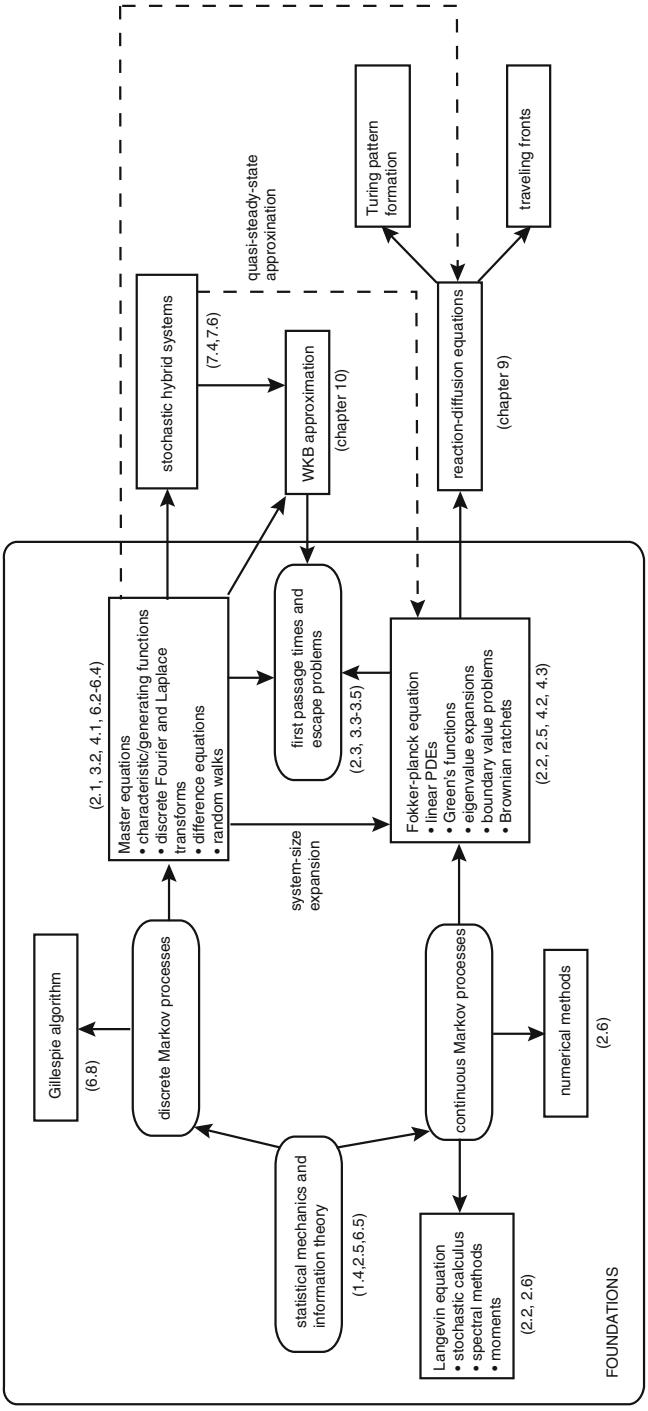


Fig. 1.12: Organizational flow chart of the book contents based primarily on mathematical themes

Part I
Foundations

Chapter 2
Diffusion in Cells: Random Walks and Brownian Motion

When one first encounters the concept of diffusion, it is usually within the context of a conservation law describing the flux of many particles moving from regions of high concentration to regions of low concentration at a rate that depends on the local concentration gradient (Fick's law). However, there are some limitations of the standard macroscopic derivation of the diffusion equation. First, it does not take into account microscopic features of the environment within which the particles diffuse. This is crucial when considering diffusive processes within a cell, since the interior of the cell is highly heterogeneous (see Box 1A). The same applies to surface diffusion within the plasma membrane. Second, with the use of advanced imaging techniques such as SPT (Sect. 1.2), it is possible to observe the movement of individual molecules, which is highly stochastic, whereas the classical diffusion equation describes the collective motion of many particles and is deterministic.

In this chapter, we consider two different microscopic theories of diffusion: random walks and overdamped Brownian motion. Both approaches will be used to model diffusion within the complex cellular environment in Chap. 7. We begin by considering a discrete random walk on a 1D lattice, which is a simple example of a discrete Markov process (Sect. 2.1). The probability distribution specifying the likelihood that the walker is at a particular lattice site after n time steps evolves according to a master equation. We show how the master equation can be solved using discrete Fourier and Laplace transforms, which in probability theory are known as characteristic functions and generating functions, respectively. The resulting solution is given by a binomial distribution, which reduces to a Gaussian distribution in an appropriate continuum limit; the latter is the fundamental solution of the diffusion equation. Background material on Laplace and Fourier transforms, and their discrete analogs, is also provided. Random walk models and various generalizations will later be used to model a variety of cellular processes, including molecular motors, polymerization of cytoskeletal filaments (Chap. 4), and anomalous diffusion (Chap. 7).

We then consider an alternative microscopic theory of diffusion based on an overdamped Brownian particle moving in a fluid-like environment (such as the cytoplasm of a cell), which is modeled in terms of a Langevin equation or SDE

P.C. Bressloff, *Stochastic Processes in Cell Biology*, Interdisciplinary Applied Mathematics 41, DOI 10.1007/978-3-319-08488-6_2

(Sect. 2.2). The latter describes the motion of the particle subject to a combination of external forces and a fluctuating force that is due to collisions with molecules in the surrounding fluid; the fluctuating force is idealized as a Wiener process. Solutions of the Langevin equation represent random sample paths or trajectories of the particle. We show how the probability density on the space of sample paths evolves according to a Fokker–Planck (FP) equation, which is a generalization of the diffusion equation applicable at the single-particle level. Other topics include the distinction between additive and multiplicative noise, Ito vs. Stratonovich interpretations of continuous stochastic processes, power spectra, and correlations. Note that continuous stochastic processes and the FP equation will appear in many chapters of the book. For in addition to describing diffusive-like motion of microscopic particles in solution, it also frequently appears in diffusion approximations of discrete Markov processes, where the continuous variable now represents the fraction of open ion channels (Chap. 3), say, or the concentration of a gene product (Chap. 6). A large-dimensional FP equation will be used to describe stochastic reaction–diffusion systems in Chap. 9 and applied to self-organizing phenomena such as cell polarization.

In Sect. 2.3 we introduce one of the most important characteristics of a diffusion process, namely, the FPT to reach a given target or boundary. This is then used to calculate the Smoluchowski reaction rate formula for diffusion-limited reactions, under the assumption that when reacting molecules come within a certain distance of each other they react immediately (Sect. 2.4). In Sect. 2.5 we tackle the general problem of diffusion in bounded domains (boundary value problems). Here we introduce some basic methods in the analysis of linear PDEs, including separation of variables and transform methods, eigenfunction expansions, and Green's functions. Finally, in Sect. 2.6 we give an informal introduction to stochastic calculus and numerical methods for simulating continuous stochastic processes.

2.1 Discrete-Time Random Walk

Consider a particle that hops at discrete times between neighboring sites on a one-dimensional (1D) lattice with unit spacing [289, 651] (see Fig. 2.1). At each step, the random walker moves a unit distance to the right with probability p or to the left with probability $q = 1 - p$. Let $P_N(r)$ denote the probability that the particle is at site r at the Nth time step. The evolution of the probability distribution is described by the discrete-time master equation

$$P_N(r) = pP_{N-1}(r-1) + qP_{N-1}(r+1), \quad r \in \mathbb{Z}, \quad N \geq 1. \tag{2.1.1}$$

If $q = p = 1/2$, then the random walk is symmetric or unbiased, whereas for $p > q$ ($p < q$) it is biased to the right (left). We will analyze this equation using transform methods, since these can be generalized to more complex random walk models such as continuous-time random walks (see Sect. 7.1.3). An introduction to continuous

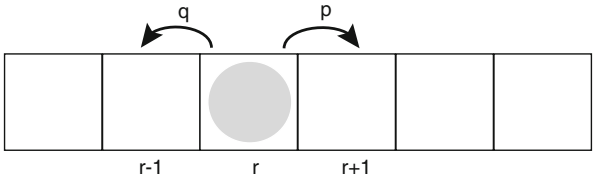


Fig. 2.1: A random walk on a 1D lattice

and discrete transform methods can be found in Box 2A. The first step is to introduce the *characteristic function* (discrete Fourier transform) for fixed N:

$$G_N(k) = \sum_{r=-\infty}^{\infty} \mathrm{e}^{ikr} P_N(r), \quad k \in [-\pi, \pi]. \tag{2.1.2}$$

The characteristic function generates moments of the random displacement variable r according to

$$\left(-i\frac{d}{dk}\right)^m G_N(k)\bigg|_{k=0} = \sum_{r=-\infty}^{\infty} r^m P_N(r) = \langle r^m \rangle, \tag{2.1.3}$$

where $\langle r^m \rangle$ is the mth order moment of r. Multiplying both sides of the master equation by e^{ikr} and summing over r gives

$$G_N(k) = (p\mathrm{e}^{ik} + q\mathrm{e}^{-ik}) G_{N-1}(k).$$

Assuming that the particle starts at the origin, $P_0(r) = \delta_{r,0}$ and $G_0(k) = 1$, we have

$$G_N(k) = u(k)^N \quad u(k) = p\mathrm{e}^{ik} + q\mathrm{e}^{-ik}.$$

Here $u(k)$ is the discrete Fourier transform of the single-step hopping probability. Finally, taking the inverse Fourier transform,

$$\begin{aligned} P_N(r) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{-ikr} u(k)^N dk \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{-ikr} \sum_{m=0}^{N} \binom{N}{m} p^m q^{N-m} \mathrm{e}^{-ik(N-2m)} dk \\ &= \frac{N!}{\left(\frac{N+r}{2}\right)!\left(\frac{N-r}{2}\right)!} p^{(N+r)/2} q^{(N-r)/2} \end{aligned} \tag{2.1.4}$$

when $N + r$ is an even integer and zero otherwise. We have used the result (see Box 2A)

$$\int_{-\pi}^{\pi} \mathrm{e}^{-ik(N-2m+r)} \frac{dk}{2\pi} = \delta_{N+r,2m}.$$

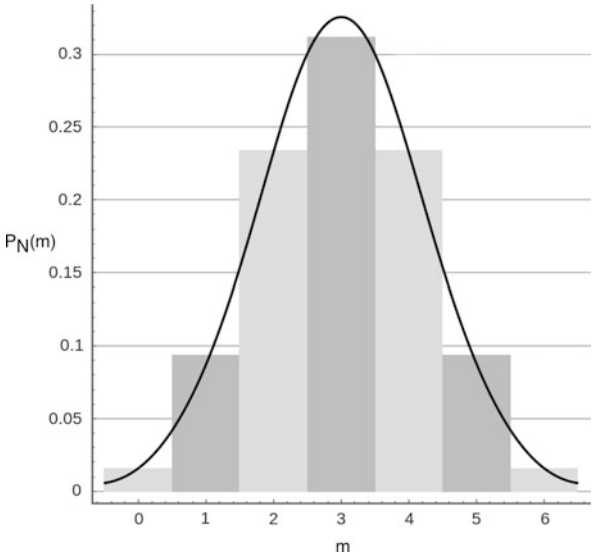


Fig. 2.2: Binomial distribution for $N = 6$ and $p = q = 1/2$. Also shown is a Gaussian fit of the binomial distribution

The distribution (2.1.4) is known as the *binomial distribution* (see Fig. 2.2). In the unbiased case $p = q = 1/2$, it gives the probability of a total of r heads in tossing a fair coin N times and is known as the *Bernoulli distribution*.

Evaluating $\log P_N(r)$ for large N using Stirling's approximation

$$\log N! \approx N \log N - N + \frac{1}{2}\log(2\pi N), \tag{2.1.5}$$

and assuming $p, q \approx 1/2$, one finds that (see Ex. 2.1 for the unbiased case)

$$P_N(r) \sim \frac{1}{\sqrt{2\pi N}} \mathrm{e}^{-[r-N(p-q)]^2/2N}. \tag{2.1.6}$$

Indeed, the Gaussian form of $P_N(r)$ in the long-time limit arises universally whenever the mean and variance of the displacement $\Delta r = r - r'$ in a single step are finite, that is,

$$\langle \Delta r \rangle = \sum_{\Delta r} \Delta r p(\Delta r) < \infty, \quad \langle \Delta r^2 \rangle = \sum_{\Delta r} (\Delta r)^2 p(\Delta r) < \infty,$$

where $p(\Delta r)$ is the probability of a step of length Δr. In the standard 1D random walk, $\Delta r = \pm 1$ and $p(1) = p, p(-1) = q$. One way to see this is to note that $u(k)$ has the small-k series expansion

$$\begin{aligned} u(k) &= \sum_{\Delta r} \mathrm{e}^{ik\Delta r} p(\Delta r) \\ &= 1 + ik\langle \Delta r\rangle - \frac{1}{2}k^2\langle \Delta r^2\rangle + \ldots \\ &\sim \mathrm{e}^{ik\langle \Delta r\rangle - \frac{1}{2}k^2\langle [\Delta r - \langle \Delta r\rangle]^2\rangle}. \end{aligned}$$

Substituting this approximation into the first line of equation (2.1.4) using the fact that the integral is dominated by the behavior in the region around $k = 0$ when N is large, the resulting Gaussian integral yields the approximation

$$P_N(r) \sim \frac{1}{\sqrt{2\pi N\sigma^2}} \mathrm{e}^{-(r - N\langle \Delta r\rangle)^2/2N\sigma^2}, \tag{2.1.7}$$

with $\sigma^2 = \langle \Delta r^2\rangle - \langle \Delta r\rangle^2$. This result is a consequence of the central limit theorem [242] (see also Sect. 1.3).

Another useful quantity when analyzing random walks is the *generating function* (discrete Laplace transform or one-sided z-transform):

$$\Gamma(r,z) = \sum_{N=0}^{\infty} z^N P_N(r). \tag{2.1.8}$$

It is often simpler to evaluate the generating function in Fourier space,

$$\hat{\Gamma}(k,z) \equiv \sum_{r=-\infty}^{\infty} \mathrm{e}^{ikr}\Gamma(r,z) = \sum_{N=0}^{\infty} z^N G_N(k),$$

assuming that we can reverse the order of summations. Since $G_N(k) = u(k)^N$, we can sum the resulting geometric series to obtain the result

$$\hat{\Gamma}(k,z) = \frac{1}{1 - zu(k)}.$$

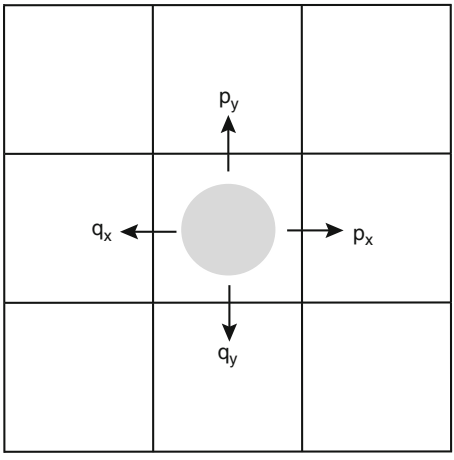


Fig. 2.3: A random walk on a 2D square lattice with $p_x + q_x + p_y + q_y = 1$

The generating function is thus given by the inverse Fourier transform

$$\Gamma(r,z) = \int_{-\pi}^{\pi} \frac{\mathrm{e}^{-ikr}}{1 - zu(k)} \frac{dk}{2\pi}.$$

It can be shown (see Ex. 2.2) that for $r = 0$ and $p = q = 1/2$ (unbiased random walk),

$$\Gamma(0,z) = (1 - z^2)^{-1/2}.$$

One immediate consequence of this result is that an unbiased 1D random walk is *recurrent*, which means that the walker is certain to return to the origin; a random walk is said to be *transient* if the probability of returning to the origin is less than one. Recurrence follows from the observation that $\Gamma(0,1) = \sum_{N=0}^{\infty} P_N(0)$ is the mean number of times that the walker visits the origin, and

$$\lim_{z \to 1^-} \Gamma(0,z) = \infty$$

for the 1D random walk. Interestingly, although the 1D random walk is recurrent, the mean time to return to the origin for the first time is infinite. This result can also be established using transform methods and generating functions (see Ex. 2.11). An unbiased random walk in 2D is also recurrent, but in 3D it is transient. An example of a 2D random walk is illustrated in Fig. 2.3. Finally, note that discrete random walks have also been used to describe the coiling of flexible polymers [53, 295] (see Sect. 4.5).

Box 2A. Transform methods.

Throughout this book we will make extensive use of transform methods, in particular, Laplace and Fourier integral transforms and their discrete analogs. Here we provide a basic introduction to such methods (see also [395]).

Laplace transforms. Let $u(t)$ be a piecewise continuous function that is of exponential order, that is,

$$u(t) \leq c\mathrm{e}^{at}, \text{ as } t \to \infty,$$

for constants $a, c > 0$. The *Laplace transform* of u is defined by

$$\mathscr{L}u(s) \equiv \tilde{u}(s) = \int_0^{\infty} u(t)\mathrm{e}^{-st} ds, \tag{2.1.9}$$

and one often writes $\mathscr{L}u = \tilde{u}$. The Laplace transform operator $\mathscr{L}$ is linear, since

$$\mathscr{L}(c_1 u_1 + c_2 u_2) = c_1 \mathscr{L}u_1 + c_2 \mathscr{L}u_2$$

for constants c_1, c_2. One of the important features of the Laplace transform (and the Fourier transform) is that it converts differential operations

in the time domain into multiplication operations in the transform domain. For example, setting $u' = du/dt$ etc.,

$$\mathscr{L}u'(s) = s\tilde{u}(s) - u(0) \tag{2.1.10a}$$

$$\mathscr{L}u''(s) = s^2\tilde{u}(s) - su(0) - u'(0), \tag{2.1.10b}$$

which can be proved using integration by parts. It follows that Laplace transforming an ordinary differential equation for $u(t)$ yields an algebraic equation for $\tilde{u}(s)$. The most difficult step, once one has solved the algebraic equation, is to find the inverse Laplace transform to recover $u(t)$. The general formula for the inverse transform requires knowledge of contour integration and takes the form

$$u(t) = \mathscr{L}^{-1}\tilde{u}(t) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} \tilde{u}(s)\mathrm{e}^{st}\,ds. \tag{2.1.11}$$

The complex contour integral is taken over the infinite vertical line (the Bromwich path) in the complex plane from $a - i\infty$ to $a + i\infty$. The real number a is chosen so that the Bromwich path lies to the right of any singularities (poles, branch points and cuts, essential points) of the function $\tilde{u}(s)$. The evaluation of the contour integral is often difficult. However, many of the Laplace transforms encountered in this book can be found in Table 2.1. One additional useful property of Laplace transforms is expressed by the *convolution theorem*.

Theorem 2.1. *Let u and v be piecewise continuous for $t \geq 0$ and of exponential order. Then*

$$\mathscr{L}(u * v)(s) = \tilde{u}(s)\tilde{v}(s), \tag{2.1.12}$$

where

$$u * v(t) \equiv \int_0^t u(t-y)v(y)dy \tag{2.1.13}$$

is the convolution of u and v. It immediately follows that $\mathscr{L}^{-1}(\tilde{u}\tilde{v}) = u * v$.

In the case of a discrete-time linear process, we can use a discrete version of the Laplace transform (also known as a one-sided z-transform)

$$\tilde{u}(z) = \sum_{n=0}^{\infty} z^n u_n. \tag{2.1.14}$$

Applying this to the first-order difference equation $u_n = au_{n-1}$ for $n \geq 1$ yields

$$\tilde{u}(z) = az\tilde{u}(z) + u_0 \implies \tilde{u}(z) = \frac{u_0}{1-az} = u_0 \sum_{n=0}^{\infty} a^n z^n.$$

The series converges provided that $|az| < 1$, in which case we immediately see that $u_n = a^n u_0$. More generally, the inverse z-transform is given by the complex integral around a closed contour C around the origin in the z-plane that does not contain any singularities of $\tilde{u}(z)$:

$$u_n = \oint_C \frac{\tilde{u}(z)}{z^{n+1}} \frac{dz}{2\pi i}. \tag{2.1.15}$$

However, one often avoids using contour integration by simply Taylor expanding the z-transform in powers of z and reading off the coefficient of z^n, as in the above example.

Fourier transforms. The *Fourier transform* of a function of one variable $u(x)$, $x \in \mathbb{R}$, is defined by the equation

$$\mathscr{F}u(k) \equiv \hat{u}(k) = \int_{-\infty}^{\infty} u(x) \mathrm{e}^{ikx} dx. \tag{2.1.16}$$

The corresponding inverse Fourier transform is

$$\mathscr{F}^{-1}\hat{u}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{u}(k) \mathrm{e}^{-ikx} dk. \tag{2.1.17}$$

An important issue is to determine the set of functions for which the Fourier transform (and its inverse) is well defined. For example, if u is integrable on $\mathbb{R}$ so that $\int_{-\infty}^{\infty} |u(x)| dx < \infty$, then

$$|\hat{u}(k)| = \left| \int_{-\infty}^{\infty} u(x) \mathrm{e}^{ikx} dx \right| \leq \int_{-\infty}^{\infty} |u(x)| dx < \infty,$$

and $\hat{u}$ exists. However, the latter may itself not be integrable. Therefore, in the application of Fourier transforms, it is common to restrict u to a much smaller class of functions such as the space of square-integrable functions denoted by $L^2(\mathbb{R})$. A few important properties of the Fourier transform are as follow. First, it converts derivatives into algebraic expressions, that is,

$$\mathscr{F}u^{(n)}(k) = (-ik)^n \hat{u}(k), \tag{2.1.18}$$

where $u^{(n)}$ denotes the nth derivative of u, and assuming that u, and its derivatives are continuous and integrable. There also exists a convolution theorem.

Theorem 2.2. *If u and v are in $L^2(\mathbb{R})$, then $u * v \in L^2(\mathbb{R})$ and*

$$\mathscr{F}(u * v)(k) = \hat{u}(k)\hat{v}(k), \tag{2.1.19}$$

where

$$(u * v)(x) \equiv \int_{-\infty}^{\infty} u(x - y) v(y) dy. \tag{2.1.20}$$

Proof. The theorem is established by interchanging the order of integration:

$$\begin{aligned}\mathscr{F}(u*v)(k) &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}u(x-y)v(y)dy\right)\mathrm{e}^{ikx}dx\\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}u(x-y)v(y)\mathrm{e}^{ikx}dx\right)dy\\ &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}u(r)v(y)\mathrm{e}^{ikr}\mathrm{e}^{iky}dr\right)dy\\ &= \int_{-\infty}^{\infty}u(r)\mathrm{e}^{ikr}dr\int_{-\infty}^{\infty}v(y)\mathrm{e}^{iky}dy = \hat{u}(k)\hat{v}(k).\end{aligned}$$

Yet another useful property is Parseval's theorem

$$\int_{-\infty}^{\infty}|u(x)|^2dx = \frac{1}{2\pi}\int_{-\infty}^{\infty}|\hat{u}(k)|^2dk. \tag{2.1.21}$$

Just as one can define a discrete Laplace transform for discrete-time processes, one can also introduce a discrete Fourier transform of spatial processes such as a random walk, which are defined on a discrete lattice. Therefore, suppose that u is a function on the space of integers $\mathbf{Z}$. The discrete Fourier transform of u is defined according to

$$(\mathscr{F}u)(k) \equiv \hat{u}(k) = \sum_{r=-\infty}^{\infty}u(r)\mathrm{e}^{ikr}, \tag{2.1.22}$$

where k is now restricted to the finite domain $(-\pi,\pi)$. The intuition behind this is that for $|k|>\pi$, the spatial oscillations $\cos(kr)$ and $\sin(kr)$ probe the function on spatial scales smaller than a unit lattice spacing where there is no information and are thus redundant. The inverse transform is

$$u(r) = \int_{-\pi}^{\pi}\hat{u}(k)\mathrm{e}^{-ikr}\frac{dk}{2\pi}. \tag{2.1.23}$$

This is straightforward to prove using the identities

$$\int_{-\pi}^{\pi}\mathrm{e}^{ik(r-s)}\frac{dk}{2\pi} = \frac{1}{2\pi i(r-s)}\left[\mathrm{e}^{i\pi(r-s)}-\mathrm{e}^{-i\pi(r-s)}\right] = 0 \text{ for } r\neq s,$$

and $\int_{-\pi}^{\pi}dk/2\pi = 1$. That is, substituting for $\hat{u}(k)$ in the inverse transform and revering the order of integration and summation,

$$\int_{-\pi}^{\pi} \hat{u}(k)\mathrm{e}^{-ikr}\frac{dk}{2\pi} = \int_{-\pi}^{\pi}\left(\sum_{s=-\infty}^{\infty} u(s)\mathrm{e}^{iks}\right)\mathrm{e}^{-ikr}\frac{dk}{2\pi}$$
$$= \sum_{s=-\infty}^{\infty} u(s)\int_{-\pi}^{\pi}\mathrm{e}^{ik(s-r)}\frac{dk}{2\pi} = \sum_{s=-\infty}^{\infty} u(s)\delta_{s,r} = u(r).$$

Note that the discrete Fourier transform should be distinguished from a Fourier series, which is an expansion of a periodic function of x in terms of a countable set of Fourier components. In other words, in a Fourier series k is unbounded but takes discrete values. Finally, consider a higher-dimensional square lattice with points $\boldsymbol{\ell} = n_1\mathbf{i} + n_2\mathbf{j}$. The corresponding discrete Fourier transform (for $d = 2$) is

$$(\mathscr{F}u)(\mathbf{k}) \equiv \hat{u}(\mathbf{k}) = \sum_{n_1=-\infty}^{\infty}\sum_{n_2=-\infty}^{\infty} u(\boldsymbol{\ell})\mathrm{e}^{i\mathbf{k}\cdot\boldsymbol{\ell}}, \tag{2.1.24}$$

with $\mathbf{k}$ the dual vector

$$\mathbf{k} = k_1\mathbf{i} + k_2\mathbf{j}, \quad k_1, k_2 \in (-\pi, \pi).$$

We will consider more general planar lattices (rhombic, hexagonal) and discrete Fourier transforms in Sect. 9.1.

$u(t)$	$\tilde{u}(s)$	$u(t)$	$\tilde{u}(s)$
1	$s^{-1}, \quad s>0$	$f(t)\mathrm{e}^{-at}$	$\tilde{f}(s+a)$
e^{at}	$\frac{1}{s-a}, \quad s>a$	$\delta(t-a)$	$\exp(-as)$
t^n	$\frac{n!}{s^{n+1}}, \quad s>0$	$H(t-a)f(t-a)$	$\tilde{f}(s)e^{-as}$
$\sin(at), \cos(at)$	$\frac{a}{s^2+a^2}, \frac{s}{s^2+a^2} \quad s>0$	$\mathrm{erf}(\sqrt{t})$	$s^{-1}(1+s)^{-1/2}, \quad s>0$
$\sinh(at), \cosh(at)$	$\frac{a}{s^2-a^2}, \frac{s}{s^2-a^2} \quad s>\lvert a\rvert$	$t^{-1/2}\exp(-a^2/4t)$	$\sqrt{\pi/s}\exp(-a\sqrt{s}), \quad s>0$
$\mathrm{e}^{at}\sin(bt)$	$\frac{b}{(s-a)^2+b^2} \quad s>a$	$1-\mathrm{erf}(a/2\sqrt{t})$	$s^{-1}\exp(-a\sqrt{s}), \quad s>0$
$\mathrm{e}^{at}\cos(bt)$	$\frac{s-a}{(s-a)^2+b^2}, \quad s>a$	$\frac{a}{2t^{3/2}}\exp(-a^2/4t)$	$\sqrt{\pi}\exp(-a\sqrt{s}), \quad s>0$

Table 2.1: Some common Laplace transforms

2.1.1 Continuum Limit of a Random Walk

Having analyzed the discrete random walk, it is now possible to take an appropriate continuum limit to obtain a diffusion equation in continuous space and time. First, introduce infinitesimal step lengths δx and δt for space and time and set $P_N(r) = \rho(x,t)\delta x$ with $x = r\delta x, t = N\delta t$. Substituting into the master equation (2.1.1) gives the following equation for the probability density $\rho(x,t)$:

$$\begin{aligned}\rho(x,t) &= p\rho(x-\delta x,t-\delta t) + q\rho(x+\delta x,t-\delta t)\\ &\approx (p+q)\left[\rho(x,t) - \frac{\partial\rho}{\partial t}\delta t\right] - (p-q)\frac{\partial\rho}{\partial x}\delta x + \frac{(p+q)}{2}\frac{\partial^2\rho}{\partial x^2}\delta x^2,\end{aligned}$$

where ρ has been Taylor expanded to first order in δt and to second order in δx. Note that $p+q=1$. Dividing through by δt and taking the continuum limit $\delta x, \delta t \to 0$ such that the quantities V, D are finite, where

$$V = \lim_{\delta x,\delta t\to 0}(p-q)\frac{\delta x}{\delta t}, \quad D = \lim_{\delta x,\delta t\to 0}\frac{\delta x^2}{2\delta t},$$

yields the advection–diffusion equation with constant drift V and diffusivity D:

$$\frac{\partial\rho(x,t)}{\partial t} = -V\frac{\partial[\rho(x,t)]}{\partial x} + D\frac{\partial^2\rho(x,t)}{\partial x^2}. \tag{2.1.25}$$

Note that $p = 0.5 + \kappa\delta x$ and $q = 0.5 - \kappa\delta x$ with $\kappa = O(1)$. For the moment, we will focus on the case of zero drift ($V = 0$), for which Eq. (2.1.25) reduces to the standard diffusion equation.

Although we have derived the diffusion equation from an unbiased random walk, it is more typically interpreted in terms of an evolution equation for a conserved quantity such as particle number rather than a probability density for a single random walker. In order to link these two interpretations, consider N noninteracting, identical diffusing particles and let $u(x,t) = Np(x,t)$. For sufficiently large N, we can treat $u(x,t)dx$ as the deterministic number of particles in the infinitesimal interval $[x, x+dx]$ at time t, with $u(x,t)$ evolving according to the diffusion equation written in the conservation form

$$\frac{\partial u}{\partial t} = -\frac{\partial J}{\partial x}, \quad J(x,t) = -D\frac{\partial u}{\partial x}, \tag{2.1.26}$$

where $J(x,t)$ is the Fickian flux of particles. Integrating the diffusion equation (2.1.26) over the interval $[x, x+dx]$ and reversing the order of integration and differentiation show that

$$\frac{d}{dt}\int_x^{x+dx} u(y,t)dy = J(x,t) - J(x+dx,t),$$

which is an expression of particle conservation. That is, the rate of change of the number of particles in $[x, x+dx]$ is equal to the net flux crossing the endpoints of the interval. Consider the initial value problem

$$\frac{\partial u}{\partial t} = D\frac{\partial u^2}{\partial x^2}, x \in \mathbb{R}, t > 0; \quad u(x,0) = f(x), x \in \mathbb{R},$$

where $f(x)$ specifies the initial data. For simplicity, we assume that $u, f \in L^2(\mathbb{R})$, that is, they are square-integrable. Taking Fourier transforms of the equation with respect to x gives

$$\frac{\partial \hat{u}(k,t)}{\partial t} = -k^2 D\hat{u}(k,t),$$

which is an ODE in t with k treated as a parameter. Its solution is

$$\hat{u}(k,t) = c(k)\mathrm{e}^{-k^2 Dt},$$

with the coefficient $c(k)$ determined by the initial data. That is, Fourier transforming the initial condition implies $\hat{u}(k,0) = \hat{f}(k)$ and, hence,

$$\hat{u}(k,t) = \hat{f}(k)\mathrm{e}^{-k^2 Dt}.$$

Applying the convolution Theorem 2.2, we have

$$u(x,t) = \int_{-\infty}^{\infty} K(x-y,t)f(y)dy,$$

where $K(x,t)$ is the inverse Fourier transform of $\mathrm{e}^{-k^2 Dt}$:

$$K(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-ikx}\mathrm{e}^{-k^2 Dt}dk = \frac{1}{\sqrt{4\pi Dt}}\mathrm{e}^{-x^2/4Dt}.$$

We thus obtain the result

$$u(x,t) = \frac{1}{\sqrt{4\pi Dt}}\int_{-\infty}^{\infty} \mathrm{e}^{-(x-y)^2/4Dt} f(y)dy. \tag{2.1.27}$$

Note the above solution still holds if we relax the requirement $f, u \in L^2(\mathbb{R})$. In particular, if we take the initial condition $f(x) = \delta(x)$, where $\delta(x)$ is the Dirac delta function (see Box 2B), then we obtain the so-called fundamental solution

$$u(x,t) = \frac{1}{\sqrt{4\pi Dt}}\mathrm{e}^{-x^2/4Dt}. \tag{2.1.28}$$

(Strictly speaking, $u(x,t)$ is a weak solution of the underlying diffusion equation [554].) Also observe that the fundamental solution corresponds to the continuum limit of the Gaussian distribution (2.1.7) for an unbiased random walk.

Box 2B. The Dirac delta function

A heuristic definition of the Dirac delta function would be that it is a "function" with the following properties:

$$\delta(0)=\infty, \quad \delta(x)=0 \text{ for all } x\neq 0, \quad \int_{\mathbb{R}}\delta(x)dx=1.$$

However, this definition is not compatible with the classical concept of a function. A rigorous definition of the Dirac delta function requires the theory of generalized functions or distributions [554]. However, an operational definition of the Dirac delta function can be constructed in terms of the limit of a sequence of Heaviside functions. Let $H(x)=1$ if $x\geq 0$ and $H(x)=0$ if $x<0$. It follows from this definition that

$$I_\varepsilon(x)\equiv\frac{H(x+\varepsilon)-H(x-\varepsilon)}{2\varepsilon}=\begin{cases}\frac{1}{2\varepsilon} & \text{if } -\varepsilon\leq x<\varepsilon\\ 0 & \text{otherwise}\end{cases}$$

It can be seen that $I_\varepsilon(x)$ has the following properties, (see Fig. 2.4):

(i) For all $\varepsilon>0$,

$$\int_{\mathbb{R}} I_\varepsilon(x)dx=\frac{1}{2\varepsilon}\times 2\varepsilon=1.$$

(ii)

$$\lim_{\varepsilon\to 0} I_\varepsilon(x)=\begin{cases}0 & \text{if } x\neq 0\\ \infty & \text{if } x=0\end{cases}$$

(iii) If $\varphi(x)$ is a smooth function that vanishes outside a bounded interval (a test function), then

$$\int_{\mathbb{R}} I_\varepsilon(x)\varphi(x)dx=\frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon}\varphi(x)dx\underset{\varepsilon\to 0}{\to}\varphi(0).$$

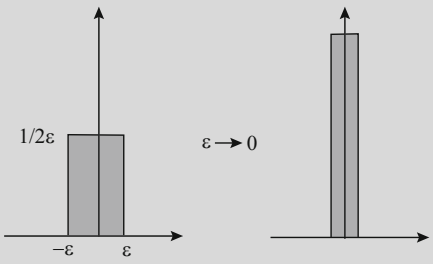


Fig. 2.4: Approximation of Dirac delta function

The third property suggests that we can define the Dirac delta function in terms of how it acts on test functions. Thus the Dirac delta function is defined as a distribution with the following properties:

$$\int_{\mathbb{R}} \delta(x)dx = 1, \quad \int_{\mathbb{R}} \delta(x)\varphi(x)dx = \varphi(0).$$

One can also introduce a shifted Dirac delta function $\delta_y(x) \equiv \delta(x-y)$,

$$\int_{\mathbb{R}} \delta_y(x)dx = 1, \quad \int_{\mathbb{R}} \delta_y(x)\varphi(x)dx = \int_{\mathbb{R}} \delta(x-y)\varphi(x)dx = \varphi(y).$$

The Heaviside construction also suggests that we can formally write $H'(x) = \delta(x)$, although again this only really makes sense in terms of test functions:

$$\int_{\mathbb{R}} H'(x)\varphi(x)dx = [H(x)\varphi(x)]_0^\infty - \int_{\mathbb{R}} H(x)\varphi'(x)dx = -\int_0^\infty \varphi'(x)dx = \varphi(0).$$

We have used integration by parts and the fact that $\varphi(x) = 0$ at $x = \infty$. Finally, note that alternative representations of the Dirac delta function include the Fourier integral,

$$\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-ikx}dk, \tag{2.1.29}$$

and the $t \to 0$ limit of the fundamental solution (2.1.28),

$$\delta(x) = \lim_{t\to 0} \frac{1}{\sqrt{4\pi Dt}}\mathrm{e}^{-x^2/4Dt}. \tag{2.1.30}$$

2.2 Continuous Stochastic Processes and the Fokker–Planck (FP) Equation

So far we have considered one approach to modeling diffusive processes, which is based on the continuum limit of a random walk; the resulting diffusion equation can be interpreted at the macroscopic level in terms of an equation for particle conservation. In this section we consider an alternative approach to modeling diffusion based on a microscopic particle moving in a fluid, such as the aqueous environment found within the interior of a cell (the cytoplasm or cytosol). The motion of the particle is modeled in terms of a continuous stochastic process evolving according to a Langevin equation or SDE. The probability density of this stochastic process satisfies a generalization of the diffusion equation known as the Fokker–Planck (FP) equation.

2.2.1 Derivation of the FP Equation from a Langevin Equation

Consider a microscopic particle such as a macromolecule moving within the cytoplasm of a cell that it is subject to some external force of size F. Collisions with fluid molecules have two distinct effects. First, they induce an apparent diffusive or Brownian motion of the particle, and second they generate an effective frictional force that opposes motion induced by the external force. In the case of microscopic particles, water acts as a highly viscous medium (low Reynolds number) so that any particle quickly approaches terminal velocity and inertial effects can be ignored (see also Box 5B). The effects of all collisions on the motion of the particle can then be represented in terms of the Langevin equation or SDE [204]

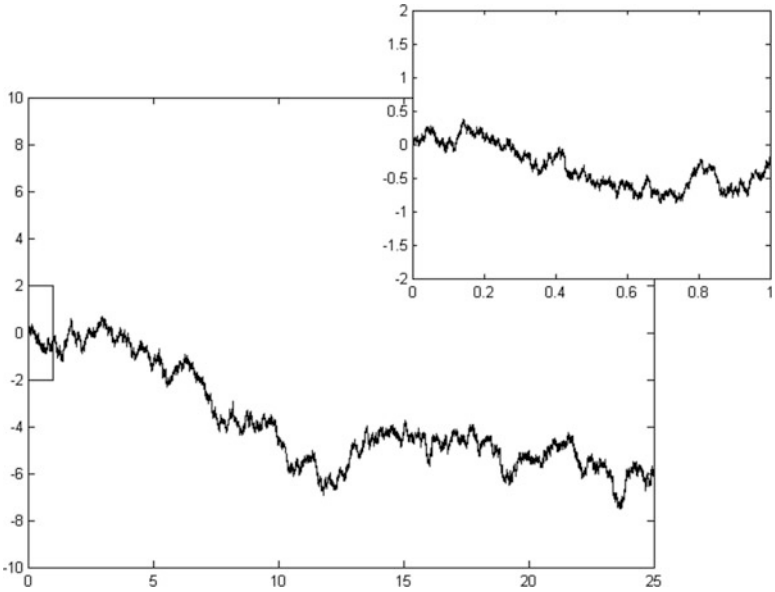


Fig. 2.5: Sample path of a Wiener process

$$dX(t) = \frac{F(X)}{\gamma}dt + \sqrt{2D}dW(t) \tag{2.2.1}$$

where $X(t)$ is the stochastic position of the particle at time t, γ is a drag coefficient, and $W(t)$ is a so-called Wiener process whose differential $dW(t)$ is a Gaussian random variable with

$$\langle dW(t)\rangle = 0, \quad \langle dW(t)dW(t')\rangle = \delta(t-t')dtdt', \tag{2.2.2}$$

where $\delta(t)$ is the Dirac delta function (Box 2B). Mathematically speaking, $W(t)$ is a continuous but everywhere non-differentiable function of time t (see Fig. 2.5). For the moment, we simply view $W(t)$ as a formal representation of the effects of a fluctuating environment. A more mathematical treatment of $W(t)$ and SDEs is presented in Sect. 2.6, together with methods for numerically simulating an SDE (see also the review by Higham [259]).

Suppose, for the moment, that F is a constant. Formally integrating Eq. (2.2.1) with $X(0) = 0$ shows that

$$X(t) = Vt + \sqrt{2D}\int_0^t dW(t')$$

with $V = F/\gamma$ the terminal velocity. Averaging with respect to the noise then implies that

$$\langle X(t)\rangle = Vt, \quad \langle (X(t) - Vt)^2\rangle = 2Dt.$$

That is,

$$\begin{aligned}\langle (X(t) - Vt)^2\rangle &= 2D\langle \int_0^t dW(t') \int_0^t dW(t'')\rangle = 2D\int_0^t \int_0^t \langle dW(t')dW(t'')\rangle \\ &= 2D\int_0^t \int_0^t \delta(t'-t'')dt'dt'' = 2D\int_0^t dt' = 2Dt.\end{aligned}$$

Hence, the MSD about the deterministic trajectory varies as $2Dt$, which suggests identifying D as a diffusion coefficient. Moreover, $X(t)$ is itself a Gaussian process whose probability density $p(x,t)$ is given by the Gaussian distribution (2.1.28) assuming the initial condition $p(x,0) = \delta(x)$. Thus, the probability density of a Brownian particle moving under the action of a constant force obeys an advection–diffusion equation of the form (2.1.25). We would like to extend this framework to the case of an x-dependent force, for which $p(x,t)$ is known to satisfy a more general Fokker–Planck (FP) equation.

We will consider a derivation of the FP equation applicable for a position-dependent force $F(x)$ along similar lines to Gardiner [204]. Since $X(t)$ is a stochastic variable, each simulation of the Langevin equation generates one sample out of the set of all possible trajectories. This motivates an alternative way of thinking about such a stochastic process, namely in terms of the conditional probability density $p(x,t|x_0,t_0)$ that the particle is at x at time t, given that it started at x_0 at time t_0. Exploiting the fact that the stochastic process is Markovian, that is, $X(t+\Delta t)$ only depends on the state at the previous time step $X(t)$, it follows that $p(x,t|x_0,t_0)$ satisfies the Chapman–Kolmogorov equation (Sect. 2.6)

$$p(x,t|x_0,t_0) = \int_{-\infty}^{\infty} p(x,t|x',t')p(x',t'|x_0,t_0)dx' \tag{2.2.3}$$

for any $t' \in [t_0, t]$. Such an equation is a defining property of a continuous Markov process. Consider an infinitesimal version of this equation by taking $t \to t + \Delta t$, $t' \to t$ and setting $w(x,t;u,\Delta t) = p(x+u, t+\Delta t|x,t)$:

$$p(x,t+\Delta t) = \int_{-\infty}^{\infty} w(x-u,t;u,\Delta t)p(x-u,t)du,$$

where the initial argument (x_0,t_0) has been suppressed. Now suppose that over a sufficiently small time window Δt, large jumps u in position are highly unlikely, so that u can be treated as a small variable (It is possible to relax this requirement - one then obtains integral terms in the evolution equation for $p(x,t)$ that represent finite jumps between states, see also **Ex. 2.3**). Taylor expanding with respect to u gives

$$p(x,t+\Delta t) = \alpha_0(x,t)p(x,t) - \partial_x[\alpha_1(x,t)p(x,t)] + \frac{1}{2}\partial_{xx}^2[\alpha_2(x,t)p(x,t)] + \ldots \tag{2.2.4}$$

where

$$\alpha_n(x,t) = \int_{-\infty}^{\infty} w(x,t;u,\Delta t)u^n du.$$

The Langevin equation (2.2.1) can be used to calculate the coefficients α_n. First, rewrite Eq. (2.2.1) in the infinitesimal form

$$X(t+\Delta t) = x + F(x)\Delta t/\gamma + \sqrt{2D}\Delta W(t),$$

given that $X(t) = x$. This implies that the transition probability w can be written as

$$\begin{aligned} w(x,t;u,\Delta t) &= \langle \delta(x+u-X(t+\Delta t))\rangle \\ &= \langle \delta(u - F(x)\Delta t/\gamma - \sqrt{2D}\Delta W(t))\rangle, \\ &= \int_{-\infty}^{\infty} \delta(u - F(x)\Delta t/\gamma - \sqrt{2D}\Delta W(t))p(\Delta W(t)) \end{aligned}$$

where p is the probability density of $\Delta W(t)$. Since

$$\Delta W(t) = \int_t^{t+\Delta t} dW(s)$$

it follows that $\Delta W(t)$ is a Gaussian random variable with zero mean and variance Δt; the corresponding probability density is

$$p(\Delta W) = \sqrt{\frac{1}{2\pi\Delta t}}\mathrm{e}^{-\Delta W^2/2\Delta t}.$$

Hence, averaging with respect to $\Delta W(t)$,

$$w(x,t;u,\Delta t) = \sqrt{\frac{1}{4\pi D\Delta t}}\mathrm{e}^{-(u-F(x)\Delta t/\gamma)^2/4D\Delta t}.$$

It follows that

$$\alpha_0 = 1, \quad \alpha_1 = F(x)\Delta t/\gamma, \quad \alpha_2 = 2D\Delta t + \alpha_1^2,$$

and $\alpha_m = \mathcal{O}(\Delta t^2)$ for $m > 2$. Substituting these results into Eq. (2.2.4) and taking the limit $\Delta t \to 0$ finally leads to the Fokker–Planck (FP) equation

$$\frac{\partial p(x,t)}{\partial t} = -\frac{1}{\gamma}\frac{\partial [F(x)p(x,t)]}{\partial x} + D\frac{\partial^2 p(x,t)}{\partial x^2}. \tag{2.2.5}$$

Note that in the limit $D \to 0$, the FP equation reduces to the so-called Liouiville equation. The latter has a general solution of the form

$$p(x,t) = \int_{\mathbb{R}} \delta(x - \phi(t,x_0))\rho(x_0)dx_0,$$

where $\phi(t,x_0)$ is the solution to the deterministic equation $\dot{x} = F(x)/\gamma$ with initial condition $x(0) = x_0$ and $\rho(x_0)$ is a probability density over initial conditions. Thus $p(x,t)$ represents a distribution of deterministic trajectories with $p(x,0) = \rho(x)$.

The 1D FP equation (2.2.5) can be rewritten as a probability conservation law according to

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial J(x,t)}{\partial x}, \tag{2.2.6}$$

where

$$J(x,t) = \frac{1}{\gamma}F(x)p(x,t) - D\frac{\partial p(x,t)}{\partial x} \tag{2.2.7}$$

is the probability flux. An equilibrium steady-state solution corresponds to the conditions $\partial p/\partial t = 0$ and $J \equiv 0$. This leads to the first-order ODE for the equilibrium density $P(x)$: $DP'(x) - \gamma^{-1}F(x)P(x) = 0$, which has the solution

$$P(x) = \mathcal{N}\mathrm{e}^{-\Phi(x)/\gamma D}.$$

Here $\Phi(x) = -\int^x F(y)dy$ is a potential energy function and $\mathcal{N}$ is a normalization factor (assuming that it exists). Comparison of the equilibrium distribution with the Boltzmann–Gibbs distribution (1.4.5) (see Sect. 1.4) yields the Einstein relation

$$D\gamma = k_B T, \tag{2.2.8}$$

where T is the temperature (in degrees Kelvin) and $k_B \approx 1.4 \times 10^{-23}\,\mathrm{JK}^{-1}$ is the Boltzmann constant. This formula relates the variance of environmental fluctuations to the strength of dissipative forces and the temperature. In the case of a sphere of radius R moving in a fluid of viscosity η, Stoke's formula can be used, that is, $\gamma = 6\pi\eta R$. For water at room temperature, $\eta \sim 10^{-3}\,\mathrm{kg\,m^{-1}\,s^{-1}}$ so that a particle of radius $R = 10^{-9}\,\mathrm{m}$ has a diffusion coefficient $D \sim 100\,\mu\mathrm{m^2\,s^{-1}}$.

It is straightforward to generalize the Langevin equation (2.2.1) to higher dimensions. Assuming for simplicity isotropic diffusion and friction, Eq. (2.2.1) becomes

$$dX_i = \frac{F_i(\mathbf{X})}{\gamma}dt + \sqrt{2D}dW_i(t), \quad i = 1,\ldots,d \tag{2.2.9}$$

with

$$\langle dW_i(t)\rangle = 0, \quad \langle dW_i(t)dW_j(t')\rangle = \delta_{i,j}\delta(t-t')dt\,dt'. \tag{2.2.10}$$

The corresponding multivariate FP equation is

$$\frac{\partial p(\mathbf{x},t)}{\partial t} = -\frac{1}{\gamma}\nabla\cdot[\mathbf{F}(\mathbf{x})p(\mathbf{x},t)] + D\nabla^2 p(\mathbf{x},t) \tag{2.2.11}$$

and the probability flux is given by the vector field

$$\mathbf{J}(\mathbf{x},t) = \frac{\mathbf{F}(\mathbf{x})}{\gamma}p(\mathbf{x},t) - D\nabla p(\mathbf{x},t). \tag{2.2.12}$$

Here ∇ denotes the gradient operator, which in Cartesian coordinates $\mathbf{x} = (x,y,z)$ (for $d = 3$) takes the form

$$\nabla = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z},$$

with $\mathbf{i}$ the unit vector in the x-direction. Similarly, ∇^2 is the Laplacian operator

$$\nabla^2 = \nabla\cdot\nabla = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

Note that the Langevin equation (2.2.1) or (2.2.9) represents diffusive-like motion from the probabilistic perspective of a single microscopic particle moving in a fluid medium. However, it is possible to reinterpret Eq. (2.2.5) or (2.2.11) as a deterministic advection-diffusion equation for the concentration $u(x,t)$ of many particles (see also Sect. 2.1.1). That is, ignoring any interactions or correlations between the particles, set $u(x,t) = Np(x,t)$ where N is the total number of particles (assumed large). Multiplying both sides of Eq. (2.2.5) by N then leads to the corresponding *Smoluchowski* equation for $u(x,t)$ with $NJ(x,t)$ interpreted as the particle flux arising from a combination of advection and Fickian diffusion. One example is the well-known Nernst–Planck equation for electrodiffusion (see Ex. 2.4). However, the relationship between macroscopic and microscopic formulations is more complicated when chemical reactions are included. Macroscopically, reactions are described in terms of the deterministic law of mass action (see Sect. 3.1), whereas microscopically they are modeled stochastically using a chemical master equation. Differences between the two levels of modeling become significant when the number of interacting molecules becomes small [651]. From the macroscopic picture of Fickian diffusion, the conservation equation $\partial_t u = -\nabla\cdot\mathbf{J}$ can lead to two different forms of the diffusion equation, depending on whether $\mathbf{J}(\mathbf{x},t) = -\nabla[D(\mathbf{x})u(\mathbf{x},t)]$ or $\mathbf{J}(\mathbf{x},t) = -D(\mathbf{x})\nabla u(\mathbf{x},t)$. (These are equivalent when D is a constant.) In order to distinguish between the two cases, it is necessary to incorporate details regarding the microscopic dynamics using, for example, kinetic theory [77]. The situation is even more complicated in anisotropic heterogeneous media, where it is no longer possible to characterize the rate of diffusion in terms of a single coefficient. One now needs to consider a diffusion tensor; see the example of active transport on

microtubular networks in Sect. 7.4. Irrespective of the particular interpretation of the FP equation (2.2.11), mathematically speaking, it is a deterministic PDE that can be analyzed using the various methods outlined in Sect. 2.5 for the diffusion equation.

2.2.2 Boundary Conditions for the FP Equation

In our study of a random walker in Sect. 2.1, we assumed that the domain was unbounded. However, diffusion of a particle within a cell is bounded and often restricted to a subcellular compartment that has a complex geometry. Therefore, it is necessary to specify the domain $\Omega \subset \mathbb{R}^d$ over which the FPE is defined and to introduce boundary conditions on $\partial\Omega$, where $\partial\Omega$ denotes the boundary of Ω. First consider the one-dimensional case ($d = 1$) with the FP equation (2.2.5) defined on the finite interval $x \in [0, L]$. This could represent the domain of a narrow ion channel (see Sect. 7.3) or a microtubular filament along which a molecular motor transports cargo (see Chap. 4). The two most common types of boundary condition at the ends $x = 0, L$ are the *Dirichlet* and *Neumann* boundary conditions. For example, at $x = 0$

$$p(0,t) = f(t) \text{ (Dirichlet) or } J(0,t) = g(t) \text{ (Neumann)}, \tag{2.2.13}$$

where $J(x,t)$ is the probability flux (2.2.7) and f, g are prescribed functions of time t, which could be time-independent. A homogeneous Dirichlet boundary condition ($f \equiv 0$) is often called an *absorbing* boundary condition, whereas a homogeneous Neumann boundary condition ($g \equiv 0$) is often called a *no-flux* or *reflecting* boundary condition. The analogous boundary conditions in higher dimensions ($d = 2, 3$) [see Eq. (2.2.11)], are

$$p(\mathbf{x},t) = f(\mathbf{x},t) \text{ (Dirichlet) or } \mathbf{J}(\mathbf{x},t)\cdot\mathbf{n}(\mathbf{x}) = g(\mathbf{x},t) \text{ (Neumann) for all } \mathbf{x} \in \partial\Omega, \tag{2.2.14}$$

where $\mathbf{n}(\mathbf{x})$ is the unit outward normal to the boundary at $\mathbf{x} \in \partial\Omega$. It is also possible to have mixed boundary conditions, in which $\partial\Omega = \partial\Omega_D \cup \partial\Omega_N$ with Dirichlet on $\partial\Omega_D$ and Neumann on $\partial\Omega_N$ (see Sect. 7.2). Alternatively, a boundary may be partially absorbing, in which case we have the *Robin* boundary condition

$$p(\mathbf{x},t) + \alpha\mathbf{J}(\mathbf{x},t)\cdot\mathbf{n}(\mathbf{x}) = 0, \quad \alpha > 0.$$

Consider the particular case of a homogeneous Neumann boundary condition. Integrating the FP equation (2.2.11) over the domain Ω, and reversing the order of integration and time differentiation, yields

$$\frac{d}{dt}\int_\Omega p(\mathbf{x},t)d\mathbf{x} = -\int_\Omega \nabla\cdot\mathbf{J}(\mathbf{x},t)d\mathbf{x} = -\int_{\partial\Omega}\mathbf{J}(\mathbf{x},t)\cdot\mathbf{n}(\mathbf{x})d\mathbf{x} = 0, \tag{2.2.15}$$

where we have used the divergence theorem [395] and imposed the boundary condition. Hence, in the case of a FPE with reflecting boundaries, the total probability

$P = \int_\Omega p(\mathbf{x},t)d\mathbf{x}$ is conserved, that is, $dP/dt = 0$, and the system typically converges to a nontrivial stationary state. On the other hand, the total probability is not conserved in the case of an absorbing boundary, which arises in FPT problems (Sect. 2.3). From the PDE perspective, there are well-established methods for solving boundary value problems for the FPE, some of which will be illustrated in Sect. 2.5 for the simpler diffusion equation.

2.2.3 The Ornstein–Uhlenbeck Process

Consider the SDE

$$dX = -kXdt + \sqrt{2D}dW(t), \tag{2.2.16}$$

where $W(t)$ is a Wiener process, and assume a fixed initial condition $X(0) = X_0$. (A Gaussian-distributed initial condition is considered in Ex. 2.5.) One way to solve this equation is to perform the change of variables $Y(t) = X(t)\mathrm{e}^{kt}$. Then

$$\begin{aligned} dY(t) &= Y(X(t+dt),t+dt) - Y(X,t) = Y(X+dX,t+dt) - Y(X,t) \\ &= (X+dX)\mathrm{e}^{k(t+dt)} - X\mathrm{e}^{kt} = kX\mathrm{e}^{kt}dt + \mathrm{e}^{kt}dX \\ &= kX\mathrm{e}^{kt}dt + \mathrm{e}^{kt}[-kXdt + \sqrt{2D}dW] = \sqrt{2D}\mathrm{e}^{kt}dW(t). \end{aligned}$$

The SDE for $Y(t)$ can now be integrated to give

$$Y(t) = Y_0 + \sqrt{2D}\int_0^t \mathrm{e}^{kt'}dW(t'),$$

that is,

$$X(t) = X_0\mathrm{e}^{-kt} + \sqrt{2D}\int_0^t \mathrm{e}^{-k(t-t')}dW(t'). \tag{2.2.17}$$

Given the explicit solution for X, we can now evaluate the mean and variance using properties of the Wiener process. First,

$$\langle X(t)\rangle = X_0\mathrm{e}^{-kt} + \sqrt{2D}\int_0^t \mathrm{e}^{-k(t-t')}\langle dW(t')\rangle = X_0\mathrm{e}^{-kt},$$

since $\langle dW\rangle = 0$. Similarly,

$$\begin{aligned} \langle [X(t) - \langle X(t)\rangle]^2\rangle &= 2D\int_0^t \mathrm{e}^{-k(t-s)}\int_0^t \mathrm{e}^{-k(t-s')}\langle dW(s)dW(s')\rangle \\ &= 2D\int_0^t \mathrm{e}^{-2k(t-s)}ds = \frac{D}{k}(1-\mathrm{e}^{-2kt}). \end{aligned} \tag{2.2.18}$$

Note that in the limit $k \to 0$, we recover the MSD of 1D Brownian motion. (The use of a change of variables to solve a Langevin equation is also considered in Ex. 2.6.)

Equation (2.2.5) implies that the FP equation for the OU process is

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial [kxp(x,t)]}{\partial x} + D\frac{\partial^2 p(x,t)}{\partial x^2}. \tag{2.2.19}$$

Taking a fixed (deterministic) initial condition $X(0) = x_0$, the initial condition of the FP equation is

$$p(x,0) = \delta(x - x_0).$$

Introduce the characteristic function (Fourier transform)

$$\Gamma(z,t) = \int_{-\infty}^{\infty} \mathrm{e}^{izx} p(x,t)dx.$$

Fourier transforming the FP equation shows that Γ satisfies the PDE (see Ex. 2.5):

$$\frac{\partial \Gamma}{\partial t} + kz\frac{\partial \Gamma}{\partial z} = -Dz^2\Gamma. \tag{2.2.20}$$

This can be solved using separation of variables (see Sect. 2.5) or the method of characteristics (see Sect. 3.6). The result is

$$\Gamma(z,t) = \exp\left[-\frac{Dz^2}{2k}(1 - \mathrm{e}^{-2kt}) + izx_0\mathrm{e}^{-kt}\right],$$

so that on applying the inverse Fourier transform (see Box 2A), we obtain the probability density

$$p(x,t) = \frac{1}{\sqrt{2\pi D[1 - \mathrm{e}^{-2kt}]/k}} e^{-(x - x_0\mathrm{e}^{-kt})^2/\left(2D[1-\mathrm{e}^{-2kt}]/k\right)}. \tag{2.2.21}$$

Note that

$$\lim_{t\to\infty} p(x,t) = p_s(x) \equiv \frac{1}{\sqrt{2\pi D/k}} e^{-kx^2/2D},$$

which is the stationary probability density.

Finally, note that the multivariate version of the OU process is given by

$$dX_i = -\sum_{j=1}^{N} A_{ij}X_j dt + \sum_{j=1}^{N} B_{ij}dW_j(t), \quad \mathbf{X}(0) = \bar{\mathbf{x}}. \tag{2.2.22}$$

The solution can be expressed formally in the matrix form (see Ex. 2.7)

$$\mathbf{X}(t) = \mathrm{e}^{-\mathbf{A}t}\bar{\mathbf{x}} + \int_0^t \mathrm{e}^{-\mathbf{A}(t-t')}\mathbf{B}d\mathbf{W}(t').$$

It can then be shown that the covariance matrix $\boldsymbol{\Sigma}(t)$ with components

$$\Sigma_{ij}(t) = \langle [X_i(t) - \langle X_i(t)\rangle][X_j(t) - \langle X_j(t)\rangle]\rangle$$

satisfies the matrix equation

$$\frac{d\boldsymbol{\Sigma}(t)}{dt} = -\mathbf{A}\boldsymbol{\Sigma}(t) - \boldsymbol{\Sigma}(t)\mathbf{A}^T + \mathbf{B}\mathbf{B}^T.$$

It follows that if $\mathbf{A}$ has distinct eigenvalues with positive real part, then $\Sigma(t) \to \Sigma_0$ where Σ_0 is the stationary covariance matrix satisfying the Riccati equation

$$\mathbf{A}\Sigma_0 + \Sigma_0 \mathbf{A}^T = \mathbf{B}\mathbf{B}^T. \tag{2.2.23}$$

The multivariate OU process will play an important role in the analysis of gene networks (see Chap. 6).

2.2.4 Multiplicative Noise

So far we have assumed that the diffusion coefficient in the Langevin equation (2.2.1) is position-independent, that is, the noise term is *additive*. The situation is considerably more involved when the term multiplying $dW(t)$ depends on $X(t)$, that is, when the noise term is *multiplicative*. The scalar Langevin equation then takes the form

$$dX(t) = A(X)dt + B(X)dW(t). \tag{2.2.24}$$

The difficulty arises since, in order to construct a solution of the SDE, we have to deal with stochastic integrals of the form $\int_0^t A(X(t))dW(t)$. Here we give a heuristic discussion of the issue—a more detailed discussion can be found in Sect. 2.6 (see also [204, 651]). Suppose for the moment that $X(t)$ and $W(t)$ are deterministic functions of time, and we can apply the theory of Riemann integration. That is, we partition the time interval $[0,T]$ into N equal intervals of size Δt with $N\Delta t = t$ and identify the value of the integral with the unique limit (assuming it exists)

$$\lim_{N\to\infty} \sum_{j=0}^{N-1} A([1-\alpha]X_j + \alpha X_{j+1})\Delta W_j$$

for $0 \le \alpha < 1$, where $\Delta W_j = W((j+1)\Delta t) - W(j\Delta t)$ and $X_j = X(j\Delta t)$. In the deterministic case, the integral is independent of α. Unfortunately, this is no longer true when we have a stochastic integral. One way to see this is to note that the ΔW_j are independent random variables. Hence, A is only statistically independent of ΔW_j when $\alpha = 0$, which is the *Ito* definition of stochastic integration. On the other hand, when $\alpha = 1/2$ we have the *Stratonovich* version. It turns out that the form of the corresponding FP equation also depends on α (see Sect. 2.6). In the Ito case,

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial [A(x)p(x,t)]}{\partial x} + \frac{1}{2}\frac{\partial^2 B(x)p(x,t)}{\partial x^2}, \tag{2.2.25}$$

whereas in the Stratonovich case

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial [A(x)p(x,t)]}{\partial x} + \frac{1}{2}\frac{\partial}{\partial x}B(x)\frac{\partial}{\partial x}[B(x)p(x,t)]. \tag{2.2.26}$$

Example 2.1. Consider the Langevin equation with linear multiplicative noise interpreted in the sense of Ito:

$$dX = X(t)a(t)dt + X(t)b(t)dW(t). \tag{2.2.27}$$

One way to solve this equation is to eliminate the multiplicative factor by performing the change of variables $Y(t) = \ln X(t)$. However, care must be taken when calculating infinitesimals, since the normal rules of calculus no longer apply for Ito stochastic variables; they do for Stratonovich variables. In particular, as shown in Sect. 2.6, one has to take into account the fact that "$dW^2 = dt$".

$$\begin{aligned} dY(t) &= \ln(X(t+dt)) - \ln X(t) = \ln(X(t) + dX(t)) - \ln X(t) = \ln(1 + dX(t)/X(t)) \\ &= \frac{dX(t)}{X(t)} - \frac{dX(t)^2}{2X(t)^2} = a(t)dt + b(t)dW(t) - \frac{1}{2}\left[a(t)dt + b(t)dW(t)\right]^2 \\ &= a(t)dt + b(t)dW(t) - \frac{b(t)^2}{2}dt + o(dt). \end{aligned}$$

Integrating this equation gives

$$Y(t) = Y_0 + \int_0^t \left[a(s) - \frac{1}{2}b(s)^2\right] ds + \int_0^t b(s)dW(s),$$

and exponentiating

$$X(t) = X_0 \exp\left(\int_0^t \left[a(s) - \frac{1}{2}b(s)^2\right] ds + \int_0^t b(s)dW(s)\right).$$

Some examples of 1D stochastic processes with multiplicative noise are considered in Exs. 2.9 and 2.10.

2.2.5 Correlations and the Power Spectrum

A very useful quantity is the *power spectrum* of a stationary stochastic process $X(t)$, which is defined as the Fourier transform of the autocorrelation function $C_X(\tau)$,

$$S_X(\omega) = \int_{-\infty}^{\infty} \mathrm{e}^{i\omega\tau} C_X(\tau)d\tau, \quad C_X(\tau) = \langle X(t)X(t+\tau)\rangle. \tag{2.2.28}$$

Consider the covariance of two frequency components of $X(t)$:

$$\begin{aligned} \langle \tilde{X}(\omega)\tilde{X}(\omega')\rangle &= \left\langle \int_{-\infty}^{\infty} \mathrm{e}^{i\omega t} X(t)dt \int_{-\infty}^{\infty} \mathrm{e}^{i\omega' t'} X(t')dt' \right\rangle \\ &= \int_{-\infty}^{\infty} \mathrm{e}^{i\omega t} \int_{-\infty}^{\infty} \mathrm{e}^{i\omega' t'} \langle X(t)X(t')\rangle dt'dt \end{aligned}$$

$$= \int_{-\infty}^{\infty} e^{i\omega t} \int_{-\infty}^{\infty} e^{i\omega' t'} \left[\int_{-\infty}^{\infty} e^{-i\Omega(t-t')} S_X(\Omega) \frac{d\Omega}{2\pi} \right] dt' dt$$
$$= \int_{-\infty}^{\infty} S_X(\Omega) \left[\int_{-\infty}^{\infty} e^{i(\omega-\Omega)t} dt \right] \left[\int_{-\infty}^{\infty} e^{i(\omega'+\Omega)t'} dt' \right] \frac{d\Omega}{2\pi},$$

assuming that it is possible to rearrange the order of integration. Using the Fourier representation of the Dirac delta function (Box 2B), $\int_{-\infty}^{\infty} e^{i\omega t} dt = 2\pi\delta(\omega)$, we have

$$\langle \tilde{X}(\omega)\tilde{X}(\omega') \rangle = \int_{-\infty}^{\infty} S_X(\Omega) \cdot 2\pi\delta(\omega-\Omega) \cdot 2\pi\delta(\omega'+\Omega) \frac{d\Omega}{2\pi}$$

which establishes a version of the Wiener–Khinchin theorem:

$$\langle \tilde{X}(\omega)\tilde{X}(\omega') \rangle = 2\pi S_X(\omega)\delta(\omega+\omega'). \tag{2.2.29}$$

The Fourier transform of a real-valued variable satisfies $\tilde{X}(-\omega) = \tilde{X}^*(\omega)$ so

$$\langle \tilde{X}(\omega)\tilde{X}^*(\omega') \rangle = 2\pi S_X(\omega)\delta(\omega-\omega'). \tag{2.2.30}$$

In the case of linear SDEs, it is possible to calculate the spectrum explicitly using the notion of a *white noise* process. Although the derivative of the Wiener process $W(t)$ does not exist, there is a sense in which the autocorrelation of the derivative does exist, which provides a useful calculational tool. For example, consider the Ornstein–Uhlenbeck process

$$dX(t) = -\kappa X(t) + dW(t).$$

Suppose that we formally rewrite this equation in terms of derivatives according to

$$\frac{dX}{dt} + \kappa X = \xi(t), \tag{2.2.31}$$

where

$$\langle \xi(t) \rangle = 0, \quad \langle \xi(t)\xi(t') \rangle = \delta(t-t').$$

The term $\xi(t)$ is known as Gaussian white noise. In order to have a stationary OU process, we take the initial time to be at $t = -\infty$. The solution can be expressed formally in terms of the integral solution

$$X(t) = \int_{-\infty}^{\infty} G(\tau)\xi(t-\tau)d\tau, \tag{2.2.32}$$

where $G(\tau)$ is known as the causal Green's function or linear response function with the important property that $G(\tau) = 0$ for $\tau < 0$. In the case of the OU process

$$G(\tau) = e^{-\tau\kappa} H(\tau),$$

where $H(t)$ is the Heaviside function. The main point to emphasize is that although $\xi(t)$ is not a mathematically well-defined object, one still obtains correct

answers when taking expectations. For example, it is clear that in the stationary state $\langle X(t)\rangle = 0$ and (for $s > 0$)

$$\begin{aligned}\langle X(t)X(t+s)\rangle &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(\tau)G(\tau')\langle \xi(t-\tau)\xi(t+s-\tau')\rangle d\tau d\tau' \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(\tau)G(\tau')\delta(s+\tau-\tau')d\tau d\tau' = \int_{-\infty}^{\infty} G(\tau)G(\tau+s)d\tau \\ &= \int_{0}^{\infty} \mathrm{e}^{-\kappa(2\tau+s)}d\tau = \frac{1}{2\kappa}\mathrm{e}^{-\kappa s}.\end{aligned}$$

This is the expected result for the autocorrelation function of the OU process.

One of the useful features of formally expressing a solution to a linear SDE in the form (2.2.32) is that one can view the dynamical system as acting as a filter of the white noise process. Applying the Wiener–Khinchin theorem to the white noise autocorrelation function, we see that the spectrum is given by the Fourier transform of a Dirac delta function, which is unity. However, once the noise has been passed through a filter with linear response function $G(t)$, the spectrum is no longer flat. This follows from applying the convolution Theorem 2.2 of Box 2A to Eq. (2.2.32):

$$\tilde{X}(\omega) = \tilde{G}(\omega)\tilde{\xi}(\omega),$$

so

$$2\pi S_X(\omega)\delta(\omega-\omega') = \tilde{G}(\omega)\tilde{G}^*(\omega')\langle\tilde{\xi}(\omega)\tilde{\xi}^*(\omega')\rangle.$$

Evaluating the various Fourier transforms, we have

$$\tilde{G}(\omega) = \int_{-\infty}^{\infty}\mathrm{e}^{i\omega t}G(t)dt = \int_{0}^{\infty}\mathrm{e}^{i\omega t}\mathrm{e}^{-\kappa t}dt = \frac{1}{\kappa - i\omega}$$

and

$$\langle\tilde{\xi}(\omega)\tilde{\xi}^*(\omega')\rangle = \int_{-\infty}^{\infty}\mathrm{e}^{i\omega t}\int_{-\infty}^{\infty}\mathrm{e}^{-i\omega' t'}\langle\xi(t)\xi(t')\rangle dt'\,dt = 2\pi\delta(\omega-\omega').$$

Hence,

$$S_X(\omega) = \frac{1}{\kappa^2+\omega^2}. \tag{2.2.33}$$

The spectrum can be used to recover the variance by noting that

$$\langle X(t)^2\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\langle\tilde{X}(\omega)\tilde{X}(\omega')\rangle\mathrm{e}^{i\omega t}\mathrm{e}^{i\omega' t}\frac{d\omega}{2\pi}\frac{d\omega'}{2\pi} = \int_{-\infty}^{\infty}S_X(\omega)\frac{d\omega}{2\pi}.$$

Substituting for $S_X(\omega)$ and using the identity

$$\int_{-\infty}^{\infty}\frac{d\omega}{\omega^2+\kappa^2} = \frac{\pi}{\kappa},$$

we see that

$$\langle X(t)^2 \rangle = \frac{1}{2\kappa}.$$

Finally, note that the formal method based on white noise is only applicable to SDEs with additive noise. In the multiplicative case, one has to use stochastic calculus (see Sect. 2.6). In this book, we will mainly express SDEs in terms of differentials and Wiener processes, restricting the use of white noise to the analysis of spectra in linear SDEs, as in Sects. 6.4 and 9.3.

2.3 First Passage Time Density and the Backward FP Equation

One of the most important ways of quantifying the efficiency of diffusive transport is in terms of the FPT to reach a target [204, 523]. In the case of intracellular transport, such a target could represent a substrate for a subsequent biochemical reaction or an exit from some bounded domain such as a chemical synapse. Consider a particle whose position evolves according to the 1D Langevin equation (2.2.1) with motion restricted to the bounded domain $x \in [0,L]$. (The FPT problem for a random walk on a lattice is considered in Ex. 2.11.) Suppose that the corresponding FP equation (2.2.5) has a reflecting boundary condition at $x = 0$ and an absorbing boundary condition at $x = L$:

$$J(0,t) = 0, \quad p(L,t) = 0.$$

We would like to determine the stochastic time $T(y)$ for the particle to exit the right-hand boundary given that it starts at location $y \in [0,L]$ at time t. As a first step, we introduce the survival probability $\mathbb{P}(y,t)$ that the particle has not yet exited the interval at time t:

$$\mathbb{P}(y,t) = \int_0^L p(x,t|y,0)dx.$$

It follows that $\text{Prob}[T(y) \leq t] = 1 - \mathbb{P}(y,t)$ and we can define the FPT density according to

$$f(y,t) = -\frac{\partial \mathbb{P}(y,t)}{\partial t} = -\int_0^L \frac{\partial}{\partial t} p(x,t|y,0)dx.$$

Using the FP equation written in conservation form (2.2.6), we see that

$$f(y,t) = \int_0^L \frac{\partial J(x,t|y,0)}{\partial x} dx = J(L,t|y,0) - J(0,t|y,0)) = J(L,t|y,0),$$

due to the reflecting boundary condition at $x = 0$. Thus the FPT density is equal to the flux through the absorbing boundary at $x = L$. In certain simple cases, the flux can be calculated explicitly, as illustrated in Ex. 2.12. However, for more general cases, it is useful to derive explicit differential equations for moments of the FPT density, in particular, the first moment or *mean first passage time* (MFPT).

In order to derive an equation for the MFPT, it is necessary to use the so-called backward FP equation. This can be derived from the Chapman–Kolmogorov equation (2.2.3) by differentiating both sides with respect to the intermediate time t':

$$0 = \int_{-\infty}^{\infty} \partial_{t'} p(x,t|x',t') p(x',t'|x_0,t_0) dx' + \int_{-\infty}^{\infty} p(x,t|x',t') \partial_{t'} p(x',t'|x_0,t_0) dx'.$$

Using the fact that $p(x',t'|x_0,t_0)$ satisfies a forward FP equation, $\partial_{t'}[p(x',t'|x_0,t_0)]$ can be replaced by terms involving derivatives with respect to x'. Integrating by parts with respect to x' then leads to the result

$$0 = \int_{-\infty}^{\infty} \left[\partial_{t'} p(x,t|x',t') + A(x') \partial_{x'} p(x,t|x',t') + D \partial_{x'x'}^2 p(x,t|x',t') \right] p(x',t'|x_0,t_0) dx'$$

where $A(x) = F(x)/\gamma$. Since p is positive, it follows that the expression in square brackets vanishes. Using time translation invariance,

$$\partial_{t'} p(x,t|x',t') = \partial_{t'} p(x,0|x',t'-t) = -\partial_t p(x,0|x',t'-t) = -\partial_t p(x,t|x',t'),$$

then yields the backward FP equation for p:

$$\partial_t p(x,t|x',t') = A(x') \partial_{x'} p(x,t|x',t') + D \partial_{x'x'}^2 p(x,t|x',t'). \tag{2.3.1}$$

Taking $x' \to y$, $t' = 0$ and integrating with respect to x shows that $\mathbb{P}(y,t)$, and hence $f(y,t)$, also satisfies a backward FP equation:

$$\frac{\partial \mathbb{P}(y,t)}{\partial t} = A(y) \frac{\partial \mathbb{P}(y,t)}{\partial y} + D \frac{\partial^2 \mathbb{P}(y,t)}{\partial y^2}. \tag{2.3.2}$$

The MFPT $\tau(y)$ is defined according to

$$\begin{aligned} \tau(y) = \langle T(y) \rangle &\equiv \int_0^{\infty} f(y,t) t dt \\ &= -\int_0^{\infty} t \frac{\partial \mathbb{P}(y,t)}{\partial t} dt = \int_0^{\infty} \mathbb{P}(y,t) dt, \end{aligned}$$

after integration by parts. Hence, integrating both sides of Eq. (2.3.2) shows that the MFPT satisfies the ODE

$$A(y) \frac{d\tau(y)}{dy} + D \frac{d^2 \tau(y)}{dy^2} = -1. \tag{2.3.3}$$

Equation (2.3.3) is supplemented by reflecting and absorbing boundary conditions for the backward FP equation:

$$\tau'(0) = 0, \quad \tau(L) = 0.$$

It is straightforward to solve Eq. (2.3.3) by direct integration [204]. First, introduce the integration factor

$$\psi(y) = \exp\left(\frac{1}{D}\int_0^y A(y')dy'\right) = \exp\left(-V(y)/k_BT\right),$$

where $D^{-1}A(y) = (D\gamma)^{-1}F(y) = -(k_BT)^{-1}V'(y)$ and $V(y)$ is a potential energy. Equation (2.3.3) becomes

$$\frac{d}{dy}\left[\psi(y)\tau'(y)\right] = -\frac{\psi(y)}{D}$$

so that

$$\psi(y)\tau'(y) = -\frac{1}{D}\int_0^y \psi(y')dy',$$

where the boundary condition $\tau'(0) = 0$ has been used. Integrating once more with respect to y and using $\tau(L) = 0$ then gives

$$\tau(y) = \int_y^L \frac{dy'}{\psi(y')}\int_0^{y'} \frac{\psi(y'')}{D}dy''. \tag{2.3.4}$$

This formula will be the starting point for analyzing escape problem in Sect. 3.3.

In the case of pure diffusion ($A(x) = 0$), we have $\psi(y) = 1$ and $\tau(y) = (L^2 - y^2)/2D$. It follows that for any finite $L - y$, $\tau(y) \to \infty$ as $L \to \infty$. Thus, although 1D diffusion is recurrent, i.e., the particle surely reaches the origin, the average time it takes is infinite. (This can also be understood in terms of the scaling properties of the FPT density.) Now suppose that L is finite and the particle starts at the left-hand boundary. The corresponding MFPT is then $\tau = L^2/D$. Within the cytosol of cells, macromolecules such as proteins tend to have diffusivities $D < 1\,\mu\mathrm{m}^2\,\mathrm{s}^{-1}$, which is due to effects such as molecular crowding. This implies that the mean time for a diffusing particle to travel a distance $100\,\mu$m is at least 10^4 s (a few hours), whereas to travel a distance 1 mm is at least 10^6 s (10 days). Since neurons, for example, which are the largest cells in humans, have axonal and dendritic protrusions that can extend from 1 mm up to 1 m, the mean travel time due to passive diffusion becomes prohibitively large, and an active form of transport becomes essential.

It is also possible to extend the above 1D analysis to the case where the particle can exit from either end [204, 523]. It is often of interest to keep track of which end the particle exits, which leads to the concept of a splitting probability. Let $S_0(x,t)$ denote the probability that the particle exits at $x' = 0$ after time t, having started at the point x. Then

$$S_0(x,t) = -\int_t^\infty J(0,t'|x,0)dt'$$

with

$$J(0,t|x,0) = A(0)p(0,t|x,0) - D\left.\frac{\partial p(y,t|x,0)}{\partial y}\right|_{y=0}.$$

Differentiating with respect to t and using the backward FP equation (2.3.1) gives

$$\begin{aligned}\frac{\partial S_0(x,t)}{\partial t} &= J(0,t|x,0) = -\int_t^\infty \frac{\partial J(0,t'|x,0)}{\partial t'}dt' \\ &= A(x)\frac{\partial S_0(x,t)}{\partial x} + D\frac{\partial^2 S_0(x,t)}{\partial x^2}. \end{aligned} \tag{2.3.5}$$

The hitting or splitting probability that the particle exits at $x' = 0$ (rather than $x' = L$) is $\Pi_0(x) = S_0(x,0)$. Moreover, the probability that the particle exits after time t, conditioned on definitely exiting through $x' = 0$, is $\text{Prob}(T_0(x) > t) = S_0(x,t)/S_0(x,0)$, where $T_0(x)$ is the corresponding conditional FPT. Since the conditional MFPT satisfies

$$\tau_0(x) = -\int_0^\infty t\frac{\partial \text{Prob}(T_0(x) > t)}{\partial t}dt = \int_0^\infty \frac{S_0(x,t)}{S_0(x,0)}dt.$$

Equation (2.3.5) is integrated with respect to t to give

$$A(x)\frac{\partial \Pi_0(x)\tau_0(x)}{\partial x} + D\frac{\partial^2 \Pi_0(x)\tau_0(x)}{\partial x^2} = -\Pi_0(x), \tag{2.3.6}$$

with boundary conditions $\Pi_0(0)\tau_0(0) = \Pi_0(L)\tau_0(L) = 0$. Finally, taking the limit $t \to 0$ in Eq. (2.3.5) and noting that $J(0,0|x,0) = 0$ for $x \neq 0$,

$$A(x)\frac{\partial \Pi_0(x)}{\partial x} + D\frac{\partial^2 \Pi_0(x)}{\partial x^2} = 0, \tag{2.3.7}$$

with boundary conditions $\Pi_0(0) = 1, \Pi_0(L) = 0$. A similar analysis can be carried out for exit through the other end $x' = L$ such that $\Pi_0(x) + \Pi_L(x) = 1$.

The construction of the FPT density can also be extended to higher spatial dimensions. Suppose that a particle evolves according to the Langevin equation (2.2.9) in a compact domain Ω with boundary $\partial\Omega$. Suppose that at time $t = 0$ the particle is at the point $\mathbf{y} \in \Omega$ and let $T(\mathbf{y})$ denote the FPT to reach any point on the boundary $\partial\Omega$. The probability that the particle has not yet reached the boundary at time t is then

$$\mathbb{P}(\mathbf{y},t) = \int_\Omega p(\mathbf{x},t|\mathbf{y},0)d\mathbf{x},$$

where $p(\mathbf{x},t|\mathbf{y},0)$ is the solution to the multivariate FP equation (2.2.11) with an absorbing boundary condition on $\partial\Omega$. The FPT density is again $f(\mathbf{y},t) = -d\mathbb{P}(\mathbf{y},t)/dt$ which, on using Eq. (2.2.11) and the divergence theorem, can be expressed as

$$f(\mathbf{y},t) = -\int_{\partial\Omega} [-\mathbf{A}(\mathbf{x})p(\mathbf{x},t|\mathbf{y},0) + D\nabla p(\mathbf{x},t|\mathbf{y},0)]\cdot d\boldsymbol{\sigma}$$

with $\mathbf{A} = \mathbf{F}/\gamma$. Similarly, by constructing the corresponding backward FP equation, it can be shown that the MFPT satisfies the equation

$$\mathbf{A}(\mathbf{y}) \cdot \nabla \tau(\mathbf{y}) + D\nabla^2 \tau(\mathbf{y}) = -1 \tag{2.3.8}$$

with $\mathbf{n} \cdot \nabla \tau(\mathbf{y}) = 0$ for $\mathbf{y} \in \partial\Omega$. Finally, note that an analogous formulation of FPTs can be formulated for discrete Markov processes evolving according to a master equation. (The particular case of a simple birth–death process is presented in Sect. 6.6, within the context of a polymerization model of gene transcription.)

2.4 Diffusion-Limited Reaction Rates

2.4.1 Smoluchowski Reaction Rate

An important example of a FPT process arises in Smoluchowski rate theory for diffusion-controlled reactions [124, 323, 523, 536, 603]. The simplest version of the theory concerns the bimolecular reaction $A + B \to AB$ for which the concentrations evolve according to the following law of mass action (see Sect. 3.1):

$$\frac{d[AB]}{dt} = k[A][B].$$

We assume that an A molecule and a B molecule react immediately to form the complex AB when they encounter each other within a reaction radius, so that the speed of reaction k is limited by their encounter rate via diffusion. (Note that k has units of volume s^{-1}. Concentrations are typically measured in *molars M* with $1\,\mathrm{molar} = 1{,}000\,\mathrm{moles/m}^3$ and $1\,\mathrm{mole} \approx 6 \times 10^{23}$ molecules (Avogadro's number).) One can then formulate the problem as an idealized first passage process, in which one A molecule, say, is fixed and treated as the center of a spherical target domain of reaction radius a, while the B molecules diffuse and are absorbed if they hit the boundary of the target domain (see Fig. 2.6a). It is assumed that the density of the particles is sufficiently small, so that reactions with other A molecules have a negligible effect on the concentration of B molecules in a neighborhood of the target molecule. The steady-state flux to the target (if it exists) is then identified as the mean reaction rate k across many targets. Let Ω denote the target domain and $\partial\Omega$ its absorbing boundary. We then need to solve the diffusion equation for the concentration $c(\mathbf{x},t)$ of background molecules exterior to the domain Ω:

$$\frac{\partial c(\mathbf{x},t)}{\partial t} = D\nabla^2 c(\mathbf{x},t), \quad c(\mathbf{x} \in \partial\Omega, t) = 0, c(\mathbf{x},0) = c_0,$$

subject to the far-field boundary condition $c(\mathbf{x},t) = c_0$ for $\mathbf{x} \to \infty$. The flux through the target boundary is

$$J = D \int_{\partial\Omega} \nabla c \cdot d\mathbf{S}.$$

Note the sign, which is due to the fact that the flux is from the exterior to the interior of the target.

Let d denote the spatial dimension of the target. For $d > 2$, a diffusing particle is transient, which means that there is a nonzero probability of never reaching the target (see Sect. 2.1). Hence, the loss of reactants by target absorption is balanced by their resupply from infinity. It follows that there exists a steady state in which the reaction rate is finite. On the other hand, for $d \leq 2$, reactants are sure to hit the target (recurrent diffusion) and a depletion zone continuously develops around the target so that the flux and reaction rate decay monotonically to zero with respect to time. Although a reaction rate does not strictly exist, it is still useful to consider the time-dependent flux as a time-dependent reaction rate. The two-dimensional case is particularly important when considering interactions of molecules embedded in the plasma membrane of a cell or the lipid bilayer surrounding an intracellular compartment.

First consider the case of a spherical target of radius a ($d = 3$). Exploiting the radial symmetry of the problem, it is possible to set $u(r,t) = rc(r,t)$ such that the 3D diffusion equation for c reduces to a 1D diffusion equation for u [523]:

$$\frac{\partial u(r,t)}{\partial t} = D \frac{\partial^2 u(r,t)}{\partial r^2}$$

with $u(r,0) = rc_0$, $u(a,t) = 0$ and $u(r,t) = rc_0$ as $r \to \infty$. Laplace transforming this equation gives $s\tilde{u}(r,s) - rc_0 = D\tilde{u}''(r,s)$, which has the solution

$$\tilde{u}(r,s) = \frac{c_0}{s}\left[r - a\mathrm{e}^{-(r-a)\sqrt{s/D}}\right].$$

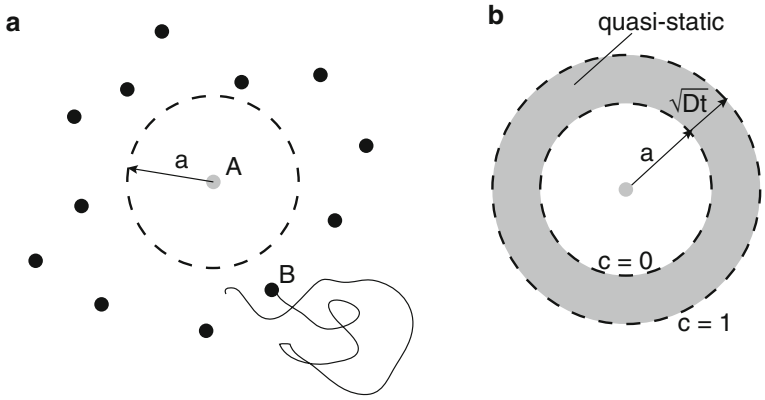


Fig. 2.6: Diffusion-limited reaction rate. (**a**) Diffusing molecules B in a neighborhood of a fixed target molecule A with reaction radius a. (**b**) Quasi-static approximation for calculating time-dependent reaction rate

Since the inverse Laplace transform of $s^{-1}[1-\mathrm{e}^{-r\sqrt{s/D}}]$ is the error function $\mathrm{erf}(r/\sqrt{4Dt})$ (see Table 2.1), where

$$\mathrm{erf}(z)=\frac{2}{\sqrt{\pi}}\int_0^z \mathrm{e}^{-r^2}dr,$$

one finds that

$$c(r,t)=c_0\left(1-\frac{a}{r}\right)+\frac{ac_0}{r}\mathrm{erf}\left[\frac{r-a}{\sqrt{4Dt}}\right].$$

It follows that the time-dependent flux is

$$\begin{aligned} J(t) &= 4\pi a^2 D\left.\frac{\partial c}{\partial r}\right|_{r=a} \\ &= 4\pi aDc_0\left(1+\frac{a}{\sqrt{\pi Dt}}\right)\underset{t\to\infty}{\to} 4\pi aDc_0. \end{aligned} \tag{2.4.1}$$

Hence, we obtain the Smoluchowski reaction rate $k=4\pi aD$. As highlighted by Redner [523], it is straightforward to generalize the steady-state result to other three-dimensional targets by making a connection with electrostatics. That is, setting $\phi(\mathbf{x})=1-c(\mathbf{x})/c_0$ in steady state, it follows that ϕ satisfies Laplace's equation with $\phi=1$ on the target boundary and $\phi=0$ at infinity, so that ϕ is equivalent to the electrostatic potential generated by a perfectly conducting object Ω held at unit potential. Moreover, the steady-state reaction rate $k=4\pi DQ$ where Q is the total charge on the surface of the conductor, which for a unit potential is equal to the capacitance, $Q=C$. Thus, determining the reaction rate for a general 3D target is equivalent to finding the capacitance of a perfect conductor with the same shape (see also [107]).

Although it is possible to calculate the exact time-dependent flux for $d\leq 2$, a much simpler method is to use a quasi-static approximation [523]. Consider, for example, a target disk of radius $r=a$. The region exterior to the disk is divided into a *near zone* that extends a distance $\sqrt{Dt}$ from the surface and a complementary *far zone* (see Fig. 2.6b). In the near zone, it is assumed that diffusing particles have sufficient time to explore the domain before being absorbed by the target so that the concentration in the near zone can be treated as almost steady or quasi-static. Conversely, it is assumed that the probability of a particle being absorbed by the target is negligible in the far zone, since a particle is unlikely to diffuse more than a distance $\sqrt{Dt}$ over a time interval of length t. Thus, $c(r)\approx c_0$ for $r>\sqrt{Dt}+a$. The near zone concentration is taken to be a radially symmetric solution of Laplace's equation, which for $d=2$ is $c(r)=A+B\log r$. Matching the solution to the boundary conditions $c(a)=0$ and $c(a+\sqrt{Dt})=c_0$ then gives (for $\sqrt{Dt}\gg a$)

$$c(r,t)\approx\frac{c_0\log(r/a)}{\log(\sqrt{Dt}/a)}.$$

The corresponding time-dependent flux is

$$J(t)\approx\frac{2\pi Dc_0}{\log(\sqrt{Dt}/a)}.$$

2.4.2 Chemoreception

In the above example, we considered a binary reaction in which the target molecule was a perfect absorber, that is, when a diffusing particle hits the target the reaction occurs immediately. One application of diffusion-limited reactions is to chemoreception. In the case of a bacterium such as *E. coli*, the cell surface is covered in receptors that detect signaling molecules in the surrounding environment (see Fig. 2.7). Treating the cell as a perfect absorber assumes that there is a sufficient number of receptors distributed on the cell surface and that binding of a signaling molecule is instantaneous when it hits the surface. There are two major simplifications of such a model—(i) receptors tend to be nonuniformly distributed on the cell surface and (ii) the rate of receptor/ligand binding $k_{\rm on}$ is finite. We shall address the second issue below. The role of receptor clustering in signal amplification will be addressed in Sect. 5.3, where we discuss the biochemical networks involved in bacterial chemotaxis.

Consider a spherical cell with M receptors distributed uniformly across its surface. Assuming that the concentration $c(r)$ of signaling molecules around the cell has reached steady state, the number of molecules absorbed per unit time is

$$\frac{dn}{dt} = Mk_{\rm on}c(a),$$

where a is the radius of the cell. From mass conservation this must be balanced by the diffusive flux through any virtual sphere of radius r, $r > a$, centered about the cell:

$$-4\pi r^2 J(r) = 4\pi r^2 D\frac{dc}{dr} = Mk_{\rm on}c(a),$$

which on integration yields

$$c(r) - c(a) = \int_a^r \frac{Mk_{\rm on}c(a)}{4\pi Dr^2}dr = \frac{Mk_{\rm on}c(a)}{4\pi Dr^2}\left(\frac{1}{a} - \frac{1}{r}\right).$$

Finally, using the far-field condition $c(\infty) = c_0$, the concentration at the surface is

$$c(a) = \frac{c_0}{1 + Mk_{\rm on}/(4\pi Da)}.$$

Hence the net absorption rate is

$$k = \frac{4\pi DaMk_{\rm on}}{4\pi Da + Mk_{\rm on}}. \tag{2.4.2}$$

In the limit $Mk_{\rm on} \to \infty$ we recover the result for a perfect absorber with $c(a) \to 0$. On the other hand, if $Mk_{\rm on} \ll Da$, then the depletion rate is so slow that $c(a) \approx c_0$, the background concentration.

Chemoreceptors allow motile *E. coli* to detect changes in concentration of a chemoattractant (food source). *E. coli* propels itself by rotating its flagella. In or-

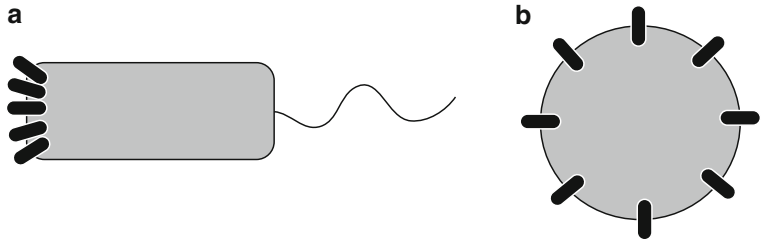


Fig. 2.7: Schematic diagram of (**a**) a polarized cell such as *E. coli* with a cluster of chemoreceptors and (**b**) a spherical cell with a uniform distribution of receptors

der to move forward, the flagella rotate together counter clockwise (CCW) enabling the bacterium to "swim" at low Reynolds number. However, when the flagella rotation abruptly changes to clockwise, the bacterium "tumbles" in place and seems incapable of going anywhere. Then the bacterium begins swimming again in some new, random direction. Swimming is more frequent as the bacterium approaches a chemoattractant (food). Tumbling, hence direction change, is more frequent as the bacterium moves away from the chemoattractant. It is the complex combination of swimming and tumbling that keeps them in areas of higher food concentrations. One important issue is why *E. coli* has to move in order to detect changes in concentration rather than simply comparing differences across its body length. The answer is that there are limitations to the sensitivity of chemoreception due to thermal noise, which means that typical concentration changes along a cell body of size 1 μm are below the signal-to-noise ratio (SNR). This observation was first made in a classical paper of Berg and Purcell [40], whose analysis will be presented in Sect. 5.1. One heuristic way to estimate the sensitivity is to assume that a bacterium integrates signals from chemoreceptors for a mean time τ_{avg}. Assuming a perfect absorber for simplicity, the total number of signaling molecules absorbed is then $N \sim aDc\tau_{\text{avg}}$. Based on the law of large numbers, we expect fluctuations in the number of molecules to vary as $\sqrt{N}$. Hence,

$$\frac{\delta c}{c} \sim \frac{\delta N}{N} \sim \frac{1}{\sqrt{Dac\tau_{\text{avg}}}}. \tag{2.4.3}$$

Taking $D \sim 10^{-5}\,\text{cm}^2/\text{s}$, $a \sim 1\,\mu\text{m}$ and a typical concentration $c = 6 \times 10^{11}$ molecules per cm^3, we have $Dac \sim 600\,\text{s}^{-1}$. Assuming that the bacterium integrates for a time $\tau_{\text{avg}} \sim 1.5\,\text{s}$, then $\delta c/c \sim 1/30$. Changes in c across 1 μm are just too small to detect. However, since the speed of motion is $v \sim 10 - 20\,\mu\text{m/s}$, it is possible to sample concentration changes of a length scale up to 30 times longer. Note that there is a limit to how large a time τ_{avg} the bacterium can integrate a chemical signal during a run, since rotational diffusion will interfere with the run's direction over longer time scales. The problem of rotational diffusion is discussed in Ex. 2.13.

2.5 Boundary Value Problems

As we have already highlighted, diffusion within the cell is bounded and often restricted to some subcellular compartment with complex geometry. This means that one has to solve an initial boundary value problem for the FPE on a bounded domain $\Omega \subset \mathbb{R}^d$ with $d = 1,2,3$. In this section we describe some methods for solving initial boundary value problems in the more specific case of the diffusion equation (see also [395, 554]). The same methods can be applied to the FP equation, although the analysis tends to be considerably more involved when the drift term and diffusivity are space-dependent [204].

2.5.1 Eigenfunction Expansions

Let $u = u(\mathbf{x},t)$ satisfy the initial boundary value problem

$$\frac{\partial u}{\partial t} = D\nabla^2 u, \quad \mathbf{x} \in \Omega, t > 0, \tag{2.5.1}$$

$$u(\mathbf{x},t) = 0, \quad \mathbf{x} \in \partial\Omega, t > 0, \quad u(\mathbf{x},0) = f(\mathbf{x}), \quad \mathbf{x} \in \Omega, \tag{2.5.2}$$

where $\partial\Omega$ denotes the boundary of Ω. For the sake of illustration, we consider the Dirichlet boundary condition $u(\mathbf{x},t) = 0, \mathbf{x} \in \partial\Omega$. However, the same methods can be applied to the Neumann or no-flux boundary condition

$$J(\mathbf{x},t) \equiv -D\mathbf{n}_{\mathbf{x}} \cdot \nabla u(\mathbf{x},t) = 0, \quad \mathbf{x} \in \partial\Omega, \tag{2.5.3}$$

where $\mathbf{n}_{\mathbf{x}}$ is the unit normal to the boundary at $\mathbf{x} \in \partial\Omega$ (with the convention that it points outward from the domain Ω). A standard method for solving this initial boundary value problem is *separation of variables*. The first step is to substitute the solution $u(\mathbf{x},t) = U(\mathbf{x})T(t)$ into the diffusion equation to give

$$U(\mathbf{x})T'(t) = DT(t)\nabla^2 U(\mathbf{x}),$$

which we rewrite as

$$\frac{T'(t)}{DT(t)} = \frac{\nabla^2 U(\mathbf{x})}{U(\mathbf{x})} = -\lambda.$$

The essential idea of the method is that λ is a constant, since it cannot be both a function of only t and only $\mathbf{x}$. It follows that we can separate the PDE into a spatial part and a temporal part according to

$$T'(t) = -\lambda DT(t) \tag{2.5.4a}$$

$$-\nabla^2 U(\mathbf{x}) = \lambda U(\mathbf{x}), \quad \mathbf{x} \in \Omega, \quad U(\mathbf{x}) = 0, \mathbf{x} \in \partial\Omega, \tag{2.5.4b}$$

where the Dirichlet boundary condition has been imposed on the spatial part.

Equation (2.5.4b) is an example of a boundary value problem for the negative Laplacian $-\nabla^2$. For each value of λ for which Eq. (2.5.4b) has a nontrivial solution $U(\mathbf{x})$, λ is called an eigenvalue and $U(\mathbf{x})$ is the corresponding eigenfunction (defined up to an arbitrary, nonzero, scalar multiplication). More formally, λ is an element of the discrete spectrum of the linear operator $-\nabla^2$ acting on the given solution domain, which is often taken to be the vector space of functions $L^2(\Omega)$. It turns out that the given Dirichlet problem has the following properties [554]:

1. The eigenvalues are real.
2. There are infinitely many eigenvalues that can be ordered as $0 < \lambda_1 \leq \lambda_2 \leq \lambda_2 \leq \ldots$ with $\lambda_n \to \infty$ as $n \to \infty$.
3. Eigenfunctions corresponding to distinct eigenvalues are orthogonal with respect to the standard inner product on Ω, that is,

$$\langle \phi_n | \phi_m \rangle \equiv \int_\Omega \phi_n(\mathbf{x}) \phi_m(\mathbf{x}) d\mathbf{x} = 0$$

   when $\lambda_n \neq \lambda_m$. The number of linearly independent eigenfunctions associated with a degenerate eigenvalue is finite, so that a Schmidt orthogonalization procedure can be used to make them orthogonal to each other, which we assume below.
4. The set of eigenfunction $\phi_n(\mathbf{x})$ is complete in the sense that any square-integrable function $F \in L^2(\Omega)$ can be uniquely represented by a generalized Fourier series

$$F(\mathbf{x}) = \sum_{n=1}^{\infty} c_n \phi_n(\mathbf{x}), \quad c_n = \frac{\langle F | \phi_n \rangle}{\|\phi_n\|^2},$$

   where c_n are the generalized Fourier coefficients and the norm is $\|\phi_n\| = \sqrt{\langle \phi_n | \phi_n \rangle}$. This means that the truncated Fourier series converges in the $L^2(\Omega)$ sense,

$$\int_\Omega \left(f(\mathbf{x}) - \sum_{n=1}^{N} c_n \phi_n(\mathbf{x}) \right)^2 d\mathbf{x} \to 0 \text{ as } N \to \infty.$$

Note that the same properties hold when the Dirichlet boundary condition is replaced by the Neumann boundary condition, except that there now exists a zero eigenvalue $\lambda_0 = 0$ whose eigenfunction $\phi_0(\mathbf{x}) =$ constant. (This reflects the fact that the diffusion equation has a nontrivial steady state in the case of a no-flux boundary condition.)

Returning to Eq. (2.5.4), we immediately see that we can identify the constant λ with one of the eigenvalues λ_n of $-\nabla^2$. Solving the equation for T then shows that we have an infinite set of solutions of the form $u_n(\mathbf{x},t) = \phi_n(\mathbf{x}) \mathrm{e}^{-D\lambda_n t}$. Since the diffusion equation is linear, we can apply the principle of superposition to write down the general solution

$$u(\mathbf{x},t) = \sum_{n=1}^{\infty} c_n \phi_n(\mathbf{x}) \mathrm{e}^{-\lambda_n D t}. \tag{2.5.5}$$

Finally, imposing the initial condition requires that

$$f(\mathbf{x}) = \sum_{n=1}^{\infty} c_n \phi_n(\mathbf{x}),$$

so that we can identify the c_n as the generalized Fourier coefficients of f. That is,

$$c_n = \frac{1}{\|\phi_n\|^2} \int_\Omega f(\mathbf{x})\phi_n(\mathbf{x}) d\mathbf{x}. \tag{2.5.6}$$

Substituting for c_n into the general solution and taking the eigenfunctions to have unit normalization ($\|\phi_n\|^2 = 1$) yields

$$u(\mathbf{x},t) = \sum_{n=1}^{\infty} \left(\int_0^L f(\mathbf{y})\phi_n(\mathbf{y}) d\mathbf{y} \right) \mathrm{e}^{-n^2\pi^2 Dt/L^2} \phi_n(\mathbf{x}).$$

Formally switching the order of summation and integration (which is valid provided that the functions are sufficiently well-behaved), the solution can be reexpressed in the compact form

$$u(\mathbf{x},t) = \int_0^L K(\mathbf{x},\mathbf{y},t) f(\mathbf{y}) d\mathbf{y}, \tag{2.5.7}$$

where

$$K(\mathbf{x},\mathbf{y},t) = \sum_{n=1}^{\infty} \mathrm{e}^{-n^2\pi^2 Dt/L^2} \phi_n(\mathbf{x})\phi_n(\mathbf{y}). \tag{2.5.8}$$

Finally, taking the limit $t \to 0$, we deduce the completeness relation

$$\sum_{n=1}^{\infty} \phi_n(\mathbf{x})\phi_n(\mathbf{y}) = \delta(\mathbf{x}-\mathbf{y}). \tag{2.5.9}$$

Example 2.2. Consider the following initial boundary value problem for the 1D diffusion equation:

$$\begin{aligned} \frac{\partial u}{\partial t} &= D\frac{\partial^2 u}{\partial x^2}, \quad 0 < x < L, t > 0. \\ u(0,t) &= 0 = u(L,t), \quad t > 0, \\ u(x,0) &= f(x), \quad 0 < x < L. \end{aligned}$$

After performing separation of variables, we obtain the eigenvalue problem

$$-U''(x) = \lambda U(x), \quad 0 < x < L, \quad U(0) = U(L) = 0.$$

The eigenvalues and eigenfunctions are thus

$$\lambda_n = \frac{n^2\pi^2}{L^2}, \quad \phi_n(x) = \sin\frac{n\pi x}{L}, \quad n = 1,2,\ldots.$$

It follows that the general solution is given by

$$u(x,t) = \sum_{n=1}^{\infty} c_n \mathrm{e}^{-n^2\pi^2 Dt/L^2} \sin\frac{n\pi x}{L}.$$

Comparison with the initial data shows that the c_n are the Fourier coefficients in the series expansion of $f(x)$,

$$f(x) = \sum_{n=1}^{\infty} c_n \sin\frac{n\pi x}{L},$$

and thus

$$c_n = \int_0^L f(\xi)\sin\frac{n\pi\xi}{L}d\xi \Big/ \int_0^L \sin^2\frac{n\pi x}{L}dx.$$

Evaluating the denominator and substituting for c_n into the general solution yields

$$u(x,t) = \frac{2}{L}\sum_{n=1}^{\infty}\left(\int_0^L f(\xi)\sin\frac{n\pi\xi}{L}d\xi\right)\mathrm{e}^{-n^2\pi^2 Dt/L^2}\sin\frac{n\pi x}{L}.$$

Formally switching the order of summation and integration (which is valid provided that the functions are sufficiently well-behaved), the solution can be reexpressed in the compact form

$$u(x,t) = \int_0^L K(x,\xi,t)f(\xi)d\xi,$$

where

$$K(x,\xi,t) = \frac{2}{L}\sum_{n=1}^{\infty}\mathrm{e}^{-n^2\pi^2 Dt/L^2}\sin\frac{n\pi\xi}{L}\sin\frac{n\pi x}{L}.$$

Some further examples of boundary value problems are considered in Exs. 2.13 and 2.14.

2.5.2 Green's Functions and Steady-State Analysis

In order for the diffusion equation to have a nontrivial steady state (time-independent solution), it is necessary to include inhomogeneous source terms in the PDE and/or inhomogeneous boundary conditions. Therefore, consider the steady-state equation in $d = 1,2,3$ dimensions

$$\nabla^2 u(\mathbf{x}) = -f(\mathbf{x}), \quad \mathbf{x}\in\Omega\subset\mathbb{R}^d, \quad u(\mathbf{x}) = g(\mathbf{x}), \mathbf{x}\in\partial\Omega. \tag{2.5.10}$$

One way to analyze inhomogeneous equations is to use Green's functions. The Green's function $G(\mathbf{x},\mathbf{y})$ for the Dirichlet boundary value problem is defined by the equation

$$\nabla_y^2 G(\mathbf{x},\mathbf{y}) = -\delta(\mathbf{x}-\mathbf{y}), \quad \mathbf{x},\mathbf{y}\in\Omega, \quad G(\mathbf{x},\mathbf{y}) = 0, \quad \mathbf{y}\in\partial\Omega, \tag{2.5.11}$$

where $\delta(\mathbf{x})$ is the Dirac delta function in $\mathbb{R}^d$. Hence, in Cartesian coordinates for $d = 3$, $\delta(\mathbf{x}) = \delta(x)\delta(y)\delta(z)$. Once one has determined the Green's function, the solution of the inhomogeneous boundary value problem can be obtained from the following Green's identity:

$$\int_\Omega \left[u(\mathbf{y})\nabla_y^2 G(\mathbf{x},\mathbf{y}) - G(\mathbf{x},\mathbf{y})\nabla_x^2 u(\mathbf{y})\right] d\mathbf{y} = \int_\Omega \nabla_y \cdot \left[u(\mathbf{y})\nabla_y G(\mathbf{x},\mathbf{y}) - G(\mathbf{x},\mathbf{y})\nabla_y u(\mathbf{y})\right] d\mathbf{y}.$$

Applying the steady-state equations to both terms on the left-hand side and using the divergence theorem on the right-hand side shows that

$$-u(\mathbf{x}) + \int_\Omega G(\mathbf{x},\mathbf{y}) f(\mathbf{y}) d\mathbf{y} = \int_{\partial\Omega} \left[u(\mathbf{y})\nabla_y G(\mathbf{x},\mathbf{y}) - G(\mathbf{x},\mathbf{y})\nabla_y u(\mathbf{y})\right] \cdot \mathbf{n} d\mathbf{y},$$

where $\mathbf{n}$ is the outward normal along the boundary $\partial\Omega$. Imposing the boundary conditions on u and G and rearranging yields the solution

$$u(\mathbf{x}) = \int_\Omega G(\mathbf{x},\mathbf{y}) f(\mathbf{y}) d\mathbf{y} - \int_{\partial\Omega} \partial_n G(\mathbf{x},\mathbf{y}) g(\mathbf{y}) d\mathbf{y}, \tag{2.5.12}$$

where $\partial_n G$ denotes the normal derivative of G.

From the properties of the spectrum of the negative Laplacian listed in Sect. 2.5.1, it follows that the Green's function has a formal expansion in terms of the complete set of orthonormal eigenfunctions:

$$G(\mathbf{x},\mathbf{y}) = \sum_{n=1}^{\infty} \frac{\phi_n(\mathbf{x})\phi_n(\mathbf{y})}{\lambda_n}. \tag{2.5.13}$$

This is straightforward to establish, since

$$\begin{aligned}\nabla_y^2 G(\mathbf{x},\mathbf{y}) &= \nabla_y^2 \left(\sum_{n=1}^{\infty} \frac{\phi_n(\mathbf{x})\phi_n(\mathbf{y})}{\lambda_n} \right) \\ &= \sum_{n=1}^{\infty} \frac{\phi_n(\mathbf{x})\nabla_y^2 \phi_n(\mathbf{y})}{\lambda_n} = -\sum_{n=1}^{\infty} \frac{\phi_n(\mathbf{x})\lambda_n \phi_n(\mathbf{y})}{\lambda_n} \\ &= -\sum_{n=1}^{\infty} \phi_n(\mathbf{x})\phi_n(\mathbf{y}) = -\delta(\mathbf{x}-\mathbf{y}).\end{aligned}$$

We have reversed the order of summation and integration and used the completeness relation (2.5.9). Note that the definition of the Green's function has to be slightly modified in the case of Neumann boundary conditions, since there exists a zero eigenvalue. The so-called generalized or modified Neumann Green's function has the eigenfunction expansion

$$G(\mathbf{x},\mathbf{y}) = \sum_{n=1}^{\infty} \frac{\phi_n(\mathbf{x})\phi_n(\mathbf{y})}{\lambda_n}. \tag{2.5.14}$$

Since the completeness relation (2.5.9) has to be extended to include the constant normalized eigenfunction $\phi_0(\mathbf{x}) = 1/\sqrt{|\Omega|}$, where $|\Omega|$ denotes the volume of the bounded domain Ω, we see that the Neumann Green's function satisfies

$$\nabla_y^2 G(\mathbf{x},\mathbf{y}) = \frac{1}{|\Omega|} - \delta(\mathbf{x}-\mathbf{y}). \tag{2.5.15}$$

One of the significant features of the Dirichlet or Neumann Green's function $G(\mathbf{x},\mathbf{y})$ in two and three dimensions is that it is singular in the limit $\mathbf{x} \to \mathbf{y}$. Moreover, these singularities take the specific form

$$G(\mathbf{x},\mathbf{y}) \sim \ln(|\mathbf{x}-\mathbf{y}) \text{ (in 2D)}\,, \quad G(\mathbf{x},\mathbf{y}) \sim \frac{1}{|\mathbf{x}-\mathbf{y}|} \text{ (in 3D)}. \tag{2.5.16}$$

The nature of these singularities is established in Box 2C.

Box 2C. The 2D and 3D Green's function for the steady-state diffusion equation.

Let us begin by considering Laplace's equation in $\mathbb{R}^2$:

$$\nabla^2 u(\mathbf{x}) = 0, \quad \mathbf{x} \in \mathbb{R}^2.$$

Since there are no boundaries, this equation is symmetric with respect to rigid body translations and rotations in the plane. This implies that if $u(\mathbf{x})$ is a solution to Laplace's equation, then so are $v(\mathbf{x}) = u(\mathbf{x}-\mathbf{a})$ and $w(\mathbf{x}) = u(\mathbf{R}_\theta \mathbf{x})$. Here $\mathbf{a}$ is a constant vector and $\mathbf{R}_\theta$ is the 2×2 rotation matrix about the origin

$$\mathbf{R}_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

This suggests that we look for a radially symmetric solution $u = u(r)$. Introducing polar coordinates, Laplace's equation becomes

$$\frac{d^2u}{dr^2} + \frac{1}{r}\frac{du}{dr} = 0, \quad 0 < r < \infty.$$

The radially symmetric solution is thus of the form

$$u(r) = C_0 \ln(r) + C_1$$

for constants C_0, C_1. Similarly, radially symmetric solutions in $\mathbb{R}^3$ satisfy Laplace's equation

$$\frac{d^2u}{dr^2} + \frac{2}{r}\frac{du}{dr} = 0, \quad 0 < r < \infty,$$

which has the solution

$$u(r) = \frac{C_0}{r} + C_1.$$

For convenience, choosing $C_1 = 0$, $C_0 = 1/4\pi$ (in 3D), and $C_0 = -1/2\pi$ (in 2D), we obtain the fundamental solution of the Laplace equation

$$K(\mathbf{x}) = -\frac{1}{2\pi}\ln|\mathbf{x}| \text{ (in 2D)}, \quad K(\mathbf{x}) = \frac{1}{4\pi|\mathbf{x}|} \text{ (in 3D)}.$$

The fundamental solution satisfies Laplace's equation everywhere except the origin, where it is singular. It turns out that K satisfies the equation

$$\nabla^2 K(\mathbf{x}) = -\delta(\mathbf{x}).$$

We will show this for the 3D case.

Let $f \in L^2(\mathbb{R}^3)$ be a function that vanishes at ∞. Define the function

$$u(\mathbf{x}) = \int_{\mathbb{R}^3} K(\mathbf{x}-\mathbf{y}) f(\mathbf{y}) d\mathbf{y} = \frac{1}{4\pi}\int_{\mathbb{R}^3} \frac{f(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|} d\mathbf{y}.$$

We will prove that $\nabla^2 u = -f$ and hence $\nabla^2 K = -\delta$. First, it is convenient to rewrite the expression for u as

$$u(\mathbf{x}) = \frac{1}{4\pi}\int_{\mathbb{R}^3} \frac{f(\mathbf{x}-\mathbf{y})}{|\mathbf{y}|} d\mathbf{y}.$$

Since $\nabla_x^2 f(\mathbf{x}-\mathbf{y}) = \nabla_y^2 f(\mathbf{x}-\mathbf{y})$,

$$\nabla^2 u(\mathbf{x}) = \int_{\mathbb{R}^3} \frac{1}{|\mathbf{y}|} \nabla_y^2 f(\mathbf{x}-\mathbf{y}) d\mathbf{y}.$$

We would like to integrate by parts, but since $\mathrm{K}(\mathbf{y})$ is singular at $\mathbf{y} = 0$, we first have to isolate the origin by surrounding it with a small sphere $B_r(0)$ of radius r. That is, we write

$$\nabla^2 u(\mathbf{x}) = \left[\int_{B_r(0)} + \int_{\mathbb{R}^3 \setminus B_r(0)}\right) \frac{1}{4\pi|\mathbf{y}|} \nabla_y^2 f(\mathbf{x}-\mathbf{y}) d\mathbf{y} \equiv I_r + J_r.$$

Here $\mathbb{R}^3 \setminus B_r(0)$ denotes $\mathbb{R}^3$ excluding the sphere around the origin. Using spherical polar coordinates,

$$\begin{aligned} |I_r| &\le \frac{\max|\nabla_y^2 f|}{4\pi} \int_{B_r(0)} \frac{1}{|\mathbf{y}|} d\mathbf{y} = \max|\nabla_y^2 f| \int_0^r \rho \, d\rho \\ &= \frac{\max|\nabla_y^2 f|}{2} r^2 \to 0 \text{ as } r \to 0. \end{aligned}$$

Recalling that f vanishes at infinity, we can integrate J_r by parts twice. First,

$$\begin{aligned} J_r &= \frac{1}{4\pi}\int_{\mathbb{R}^3\setminus B_r(0)} \left[\nabla_y \cdot \left(\frac{1}{|\mathbf{y}|}\nabla_y f(\mathbf{x}-\mathbf{y})\right) - \nabla_y \frac{1}{|\mathbf{y}|}\cdot \nabla_y f(\mathbf{x}-\mathbf{y})\right] d\mathbf{y} \\ &= \frac{1}{4\pi}\int_{\mathbb{R}^3\setminus B_r(0)} \nabla_y \cdot \left(\frac{1}{|\mathbf{y}|}\nabla_y f(\mathbf{x}-\mathbf{y}) - f(\mathbf{x}-\mathbf{y})\nabla_y \frac{1}{|\mathbf{y}|}\right) d\mathbf{y} \\ &\quad + \frac{1}{4\pi}\int_{\mathbb{R}^3\setminus B_r(0)} f(\mathbf{x}-\mathbf{y})\nabla_y^2 \frac{1}{|\mathbf{y}|} d\mathbf{y}. \end{aligned}$$

Using the fact that $\nabla_y^2(1/|\mathbf{y}|) = 0$ in $\mathbb{R}^3\setminus B_r(0)$, and applying the divergence theorem, we have

$$J_r = \frac{1}{4\pi}\int_{\partial B_r(0)} \left(\frac{1}{|\mathbf{y}|}\nabla_y f(\mathbf{x}-\mathbf{y}) - f(\mathbf{x}-\mathbf{y})\nabla_y \frac{1}{|\mathbf{y}|}\right)\cdot \mathbf{n}_y d\mathbf{y}.$$

The first integral vanishes in the limit $r \to 0$, since

$$\frac{1}{4\pi r}\left|\int_{\partial B_r(0)} \nabla_y f(\mathbf{x}-\mathbf{y})\cdot \mathbf{n}_y d\mathbf{y}\right| \leq r \max|\nabla_y f| \to 0.$$

On the other hand, since $\nabla_y(1/|\mathbf{y}|) = -\mathbf{y}/|\mathbf{y}|^3$ and $\mathbf{n}_y = -\mathbf{y}/r$, the second integral yields

$$\frac{1}{4\pi}\int_{\partial B_r(0)} f(\mathbf{x}-\mathbf{y})\nabla_y \frac{1}{|\mathbf{y}|}\cdot \mathbf{n}_y d\mathbf{y} = \frac{1}{4\pi r^2}\int_{\partial B_r(0)} f(\mathbf{x}-\mathbf{y}) d\mathbf{y} \to f(\mathbf{x}) \text{ as } r \to 0.$$

We conclude that $I_r \to 0$ and $J_r \to -f(\mathbf{x})$ as $r \to 0$, which implies that $\nabla^2 u = -f$ and $\nabla^2 K = -\delta$. A similar analysis can be carried out in 2D using the logarithmic fundamental solution. Finally, given the properties of the fundamental solution $K(\mathbf{x})$, we can construct the Green's function for a boundary value problem in terms of $K(\mathbf{x})$ and a non-singular or regular part that satisfies Laplace's equation everywhere, that is,

$$G(\mathbf{x},\mathbf{y}) = K(\mathbf{x}-\mathbf{y}) + R(\mathbf{x},\mathbf{y}).$$

We will use this result in Sect. 7.2, when considering narrow escape problems and diffusion to small targets.

2.6 Appendix: An Informal Introduction to Stochastic Calculus

In this appendix we present an informal introduction to stochastic calculus, following along the lines of Jacobs [298]. A more detailed treatment can be found in Gardiner [204], and a rigorous mathematical account can be found in [483].

The basic approach is to treat a continuous-time stochastic process as the limit of a discrete-time process. That is, an SDE prescribes how a stochastic variable $X(t)$ changes in each infinitesimal time step dt. Determining changes over finite times then requires evaluating an associated stochastic integral. In order to make sense of this, we discretize time into small, but finite, intervals of duration Δt and consider a corresponding stochastic difference equation for $X_n = X(n\Delta t)$. A more abstract formulation of probability theory and discrete-time stochastic processes is presented in Chap. 11.

2.6.1 What Is a Continuous Stochastic Process?

Suppose that an experiment is carried out over a time interval of length T and has a given set of possible outcomes Ω. In the case of tracking a single molecule diffusing in the cell membrane, Ω could be the set of all possible trajectories. On the other hand, in the case of an ion channel, Ω could specify whether the channel is open or closed. Each time the experiment is run, one obtains a particular realization of a continuous-time stochastic process (or random function) $X(\omega,t)$ with $\omega \in \Omega$. (If the time interval is sampled at discrete times, then one has a discrete-time stochastic process or random sequence.) For fixed ω, $X(\omega,t) = X_\omega(t)$ is a function of time corresponding to a particular trajectory in state space, which is specified by the parameter ω. On the other hand, fixing time t yields a family of random variables $X(\omega,t) = X_t(\omega)$ that are parameterized by t. In the case of diffusion $X_t(\omega)$ is a continuous random variable, whereas for an ion channel $X_t(\omega)$ is a discrete random variable (see Sect. 1.3). For concreteness, we will focus on the continuous case. In physical and biological applications, the explicit dependence on the events ω and the nature of the underlying probability space are ignored, and one simply writes $X = X(t)$. The cumulative distribution function of the stochastic process is defined according to

$$P(x,t) = \text{prob}[X(t) \leq x]. \tag{2.6.1}$$

Using the frequency interpretation of probability, this represents the fraction of trials for which the trajectory through state space does not exceed the value x at time t. One can then define the corresponding probability density (assuming it exists) according to

$$p(x,t) = \frac{\partial P(x,t)}{\partial x}. \tag{2.6.2}$$

Moreover, as in classical probability theory, we can introduce joint cumulative distributions and densities

$$P(x_1,\ldots,x_n;t_1,\ldots,t_n) = \text{Prob}[X(t_1) \leq x_1,\ldots,X(t_n) \leq x_n].$$

and

$$p(x_1,\ldots,x_n;t_1,\ldots,t_n) = \frac{\partial^n P(x_1,\ldots,x_n;t_1,\ldots,t_n)}{\partial x_1 \ldots \partial x_n}.$$

A further extension is the conditional probability density, which takes the form (for $n = 2$)

$$p(x_2,t_2|x_1,t_1) = \frac{p(x_1,x_2;t_1,t_2)}{p(x_1,t_1)}. \tag{2.6.3}$$

Thus, $p(x_2,t_2|x_1,t_1)$ is the probability density for $X(t_2)$ conditioned on $X(t_1) = x_1$.

Given a probability density p, one can define various moments of the stochastic process. Some important examples are the mean

$$x(t) = \langle X(t)\rangle = \int_{-\infty}^{\infty} xp(x,t)dx, \tag{2.6.4}$$

and the two-point *autocorrelation function*

$$\langle X(t_1)X(t_2)\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2p(x_1,x_2;t_1,t_2)dx_1dx_2. \tag{2.6.5}$$

A related quantity is the *covariance* given by

$$C(t_1,t_2) = \langle (X(t_1)-x(t_1))(X(t_2)-x(t_2))\rangle = \langle\langle X(t_1)X(t_2)\rangle\rangle. \tag{2.6.6}$$

The equal-time covariance $C(t,t)$ is the corresponding variance. Double brackets are often used to denote cumulants of the stochastic process. The latter are defined using a generating function:

$$\left\langle \exp\left(-i\int_0^t X(t')dt'\right)\right\rangle \tag{2.6.7}$$
$$= \exp\left[\sum_{n=1}^{\infty}\frac{(-i)^m}{m!}\int_0^t\ldots\int_0^t \langle\langle X(t_1)X(t_2)\ldots X(t_n)\rangle\rangle dt_1dt_2\ldots dt_n\right].$$

An important concept in stochastic processes is stationarity: a stochastic process $X(t)$ is *stationary* if every joint probability distribution for finite n is invariant under a global time shift:

$$P(x_1,x_2,\ldots,x_n;t_1+\tau,t_2+\tau,\ldots,t_n+\tau) = P(x_1,x_2,\ldots,x_n;t_1,t_2,\ldots,t_n)$$

for arbitrary τ. It follows that $P(x,t)$ is time-independent and the covariance $C(t.t') = C(t-t')$.

A very important type of stochastic process is a *Markov process*, which is defined by the property that the conditional probability density satisfies

$$p(x_n,t_n|x_1,\ldots,x_{n-1};t_1,\ldots,t_{n-1}) = p(x_n,t_n|x_{n-1},t_{n-1}). \tag{2.6.8}$$

In other words, given (x_{n-1},t_{n-1}), the process has no "memory" of values at earlier times. It follows that a Markov process is fully determined by the two functions $p(x_1,t_1)$ and $p(x_2,t_2|x_1,t_1)$. For example,

$$\begin{aligned} p(x_1,x_2,x_3;t_1,t_2,t_3) &= p(x_3,t_3|x_1,x_2;t_1,t_2)p(x_1,x_2;t_1,t_2) \\ &= p(x_3,t_3|x_1,x_2;t_1,t_2)p(x_2,t_2|x_1,t_1)p(x_1,t_1) \\ &= p(x_3,t_3|x_2,t_2)p(x_2,t_2|x_1,t_1)p(x_1,t_1). \end{aligned} \tag{2.6.9}$$

However, the functions $p(x_1,t_1)$ and $p(x_2,t_2|x_1,t_1)$ cannot be chosen arbitrarily, since they must obey two important identities. The first is obtained by integrating

$$p(x_1,x_2;t_1,t_2) = p(x_2,t_2|x_1,t_1)p(x_1,t_1)$$

with respect to x_1:

$$p(x_2,t_2) = \int_{-\infty}^{\infty} p(x_2,t_2|x_1,t_1)p(x_1,t_1)dx_1. \tag{2.6.10}$$

The other is obtained by integrating Eq. (2.6.9) with respect to x_2, assuming that $t_1 < t_2 < t_3$:

$$p(x_1,x_3;t_1,t_3) = p(x_1,t_1)\int_{-\infty}^{\infty} p(x_3,t_3|x_2,t_2)p(x_2,t_2|x_1,t_1)dx_2.$$

Since $p(x_1,x_3;t_1,t_3) = p(x_3,t_3|x_1,t_1)p(x_1,t_1)$, this reduces to the Chapman–Kolmogorov equation

$$p(x_3,t_3|x_1,t_1) = \int_{-\infty}^{\infty} p(x_3,t_3|x_2,t_2)p(x_2,t_2|x_1,t_1)dx_2. \tag{2.6.11}$$

2.6.2 Ito Stochastic Integrals

Suppose that the time interval $[0,T]$ is divided into N increments of size $\Delta t = T/N$ and set $t_n = n\Delta t$. Consider the stochastic difference equation

$$\Delta X(t_n) \equiv X(t_{n+1}) - X(t_n) = \Delta W_n,$$

where ΔW_n, $n = 0,\ldots,N-1$, are independent and identically distributed Gaussian variables with zero mean and variance $\sigma^2 = \Delta t$:

$$P(\Delta W) = \frac{1}{\sqrt{2\pi\Delta t}} e^{-(\Delta W)^2/2\Delta t}. \tag{2.6.12}$$

(Note that a sequence of random variables is independent and identically distributed (i.i.d.) if each random variable has the same probability distribution as the others

and all are mutually independent.) Iterating the difference equation starting from $X(0) = 0$ yields

$$X_n \equiv X(n\Delta t) = \sum_{j=0}^{n-1} \Delta W_j.$$

Using the fact that the sum of Gaussian random variables is also a Gaussian, it follows that the probability density for X_n is a Gaussian. Thus, we only need to determine its mean and variance. Since the ΔW_j are all independent, we have

$$\langle X_n \rangle = \sum_{j=0}^{n-1} \langle \Delta W_j \rangle = 0, \quad \mathrm{Var}(X_n) = \sum_{j=0}^{n-1} \mathrm{Var}(\Delta W_j) = n\Delta t,$$

and

$$P(X_n) = \frac{1}{\sqrt{2\pi n \Delta t}} \mathrm{e}^{-X_n^2/(2n\Delta t)}.$$

We can now construct a corresponding continuous-time process by taking the limit $N \to \infty$ such that $\Delta t \to 0$ with $N\Delta T = T$ fixed. In particular,

$$X(T) = \lim_{N\to\infty} \sum_{j=0}^{N-1} \Delta W_j \equiv \int_0^T dW(t) \equiv W(T),$$

where $W(T)$ is identified as a Wiener process—a Gaussian process with independent and stationary increments. (For the moment, we will not worry about the precise meaning of convergence and limits of stochastic variables—this will be addressed below.) $X(T)$ is still a Gaussian, whose mean and variance are obtained by taking the limit $N \to \infty$ of the results for X_n. We deduce that $W(t)$ has the Gaussian probability density

$$P(w(t)) = \frac{1}{\sqrt{2\pi t}} \mathrm{e}^{-w(t)^2/2t}.$$

Now consider the modified stochastic difference equation

$$X_{n+1} - X_n = f(t_n)\Delta W_n,$$

where $f(t)$ is a deterministic function of time. Once again X_n is a Gaussian random variable, with

$$\langle X_n \rangle = \sum_{j=0}^{n-1} \langle f(t_j)\Delta W_j \rangle = 0, \quad \mathrm{Var}(X_n) = \sum_{j=0}^{n-1} \mathrm{Var}(f(t_j)\Delta W_j) = \sum_{j=0}^{n-1} f(t_j)^2 \Delta t.$$

Taking the continuum limit along identical lines to the previous case yields the continuous-time Gaussian variable

$$X(T) = \lim_{N\to\infty} \sum_{j=0}^{N-1} f(t_j)\Delta W_j \equiv \int_0^T f(t) dW(t), \tag{2.6.13}$$

with zero mean and variance

$$\mathrm{Var}(X(T)) = \int_0^T f(s)^2 ds. \tag{2.6.14}$$

Substituting for $X(T)$ into this equation gives

$$\left\langle \int_0^T f(t)dW(t) \int_0^T f(s)dW(s) \right\rangle = \int_0^T f(s)^2 ds,$$

which can be captured by the rule

$$\langle dW(t)dW(s)\rangle = \delta(t-s)dt\,ds. \tag{2.6.15}$$

However, care must be taken with this rule when $\delta(t-s)$ appears inside an integral having t or s as one of its limits. For example, consider the double stochastic integral

$$\int_0^T \left[\int_0^t f(s)dW(s) \right] g(t)dW(t) \equiv \lim_{N\to\infty} \sum_{n=0}^{N-1} \left[\sum_{m=0}^{n-1} f(t_m)\Delta W_m \right] g(t_n)\Delta W_n.$$

We see that there are no terms in the double sum on the right-hand side that have a product of Wiener increments in the same time interval. Thus, taking the expectation of both sides,

$$\left\langle \int_0^T \left[\int_0^t f(s)dW(s) \right] g(t)dW(t) \right\rangle = 0.$$

Hence, we require

$$\int_0^t f(s)\delta(t-s)ds = 0, \quad \int_0^t f(s)\delta(s)ds = f(0). \tag{2.6.16}$$

Following the previous examples, let us turn to a discretized version of the general SDE for $X(t)$,

$$dX = a(X,t)dt + b(X,t)dW(t), \tag{2.6.17}$$

which takes the form

$$X_{n+1} - X_n = a(X_n,t_n)\Delta t + b(X_n,t_n)\Delta W_n. \tag{2.6.18}$$

Iterating this equation starting from a fixed $X(0) = x_0$ yields

$$X_N = x_0 + \sum_{n=0}^{N-1} a(X_n,t_n)\Delta t + \sum_{n=0}^{N-1} b(X_n,t_n)\Delta W_n.$$

The continuum limit then gives the stochastic integral equation

$$X(T) = x_0 + \int_0^T a(X(t),t)dt + \int_0^T b(X(t),t)dW(t), \tag{2.6.19}$$

with the final term defined as the Ito stochastic integral

$$\int_0^T b(X(t),t)dW(t) = \lim_{N\to\infty} \sum_{n=0}^{N-1} b(X_n,t_n)\Delta W_n. \tag{2.6.20}$$

The integral equation is not very useful for generating an explicit solution for $X(t)$. However, from the definition of the Ito stochastic integral, it immediately follows that

$$\left\langle \int_0^T b(X(t),t)dW(t) \right\rangle = 0, \tag{2.6.21}$$

since X_n is a function of previous Wiener increments $\Delta W_{n-1},\ldots,\Delta W_0$ so it is uncorrelated with ΔW_n. The stochastic difference equation (2.6.18) is the starting point for developing numerical schemes for solving an SDE. However, if one is interested in carrying out explicit calculations, it is usually more useful to go to the associated FPE for the probability density. In order to derive the FP equation from the corresponding SDE, we first need to consider the object $(dW)^2$.

In terms of Wiener increments,

$$\int_0^T (dW(t))^2 = \lim_{N\to\infty} \sum_{n=0}^{N-1} (\Delta W_n)^2.$$

Taking the expectation of both sides and using the fact that each ΔW_n is an i.i.d. gives

$$\left\langle \int_0^T (dW(t))^2 \right\rangle = \int_0^T \langle (dW(t))^2 \rangle = \int_0^T dt = T. \tag{2.6.22}$$

What about the variance? Using the Gaussian probability density (2.6.12), it is simple to show that

$$\mathrm{Var}[(\Delta W)^2] = 2(\Delta t)^2 = 2T^2/N^2.$$

Hence,

$$\begin{aligned}\mathrm{Var}\left[\int_0^T (dW(t))^2\right] &= \lim_{N\to\infty} \mathrm{Var}\left[\sum_{n=0}^{N-1} (\Delta W_n)^2\right] = \lim_{N\to\infty} \sum_{n=0}^{N-1} \mathrm{Var}\left[(\Delta W_n)^2\right] \\ &= \lim_{N\to\infty} \frac{2T^2}{N} = 0.\end{aligned}$$

We thus obtain the surprising result that the integral of $(dW)^2$ is deterministic and thus equal to its mean:

$$\int_0^T (dW(t))^2 = T = \int_0^T dt. \tag{2.6.23}$$

In other words, we can set $(dW)^2 = dt$ inside integrals, a result known as Ito's rule (see below). Using the higher moments of Gaussians, it can also be shown that $dW^m = 0$ for $m > 2$. We now consider a result from stochastic calculus, which will be useful when discussing numerical simulations, in particular Milstein's method (see Sect. 2.6.6).

Example 2.3. $\int_{t_0}^{t} W(t')dW(t')$. Consider the discrete sum

$$\begin{aligned} S_n &= \sum_{n=0}^{N-1} W_n \Delta W_n \\ &= \frac{1}{2}\sum_{n=0}^{N-1}\left[(W_n+\Delta W_n)^2 - W_n^2 - (\Delta W_n)^2\right] \\ &= \frac{1}{2}[W(t)-W(t_0)]^2 - \frac{1}{2}\sum_{n=0}^{N-1}(\Delta W_n)^2, \end{aligned}$$

where $W_0 = W(t_0)$ and $W_{N-1} = W(t)$. We now calculate the mean and variance of the last term. First,

$$\left\langle \sum_{n=0}^{N-1}(\Delta W_n)^2 \right\rangle = \sum_{n=0}^{N-1}\left\langle (\Delta W_n)^2\right\rangle = \sum_{n=0}^{N-1}(t_{n+1}-t_n) = t-t_0.$$

Second,

$$\begin{aligned} &\left\langle \left[\sum_{n=0}^{N-1}(\Delta W_n)^2 - (t-t_0)^2\right]^2 \right\rangle \\ &\quad = \left\langle \left[\sum_{n=0}^{N-1}(\Delta W_n)^4 + 2\sum_{n<m}(\Delta W_n)^2(\Delta W_m)^2 - 2(t-t_0)\sum_{n=0}^{N-1}(\Delta W_n)^2 + (t-t_0)^2\right]\right\rangle. \end{aligned}$$

Since ΔW_n and ΔW_m are independent Gaussian random variables for $n \neq m$, we have

$$\left\langle (\Delta W_n)^2(\Delta W_m)^2\right\rangle = (t_{n+1}-t_n)(t_{m+1}-t_m),$$

and the fourth moment of a Gaussian is given by

$$\left\langle (\Delta W_n)^4\right\rangle = 3\left\langle (\Delta W_n)^2\right\rangle^2 = 3(t_{n+1}-t_n)^2.$$

Hence,

$$\begin{aligned} \left\langle \sum_{n=0}^{N-1}(\Delta W_n)^2\right\rangle &= 2\sum_{n=0}^{N-1}(t_{n+1}-t_n)^2 + \sum_{n,m}[(t_{n+1}-t_n)-(t-t_0)][(t_{m+1}-t_m)-(t-t_0)] \\ &= 2\sum_{n=0}^{N-1}(t_{n+1}-t_n)^2 \to 0 \text{ as } N\to\infty. \end{aligned}$$

We deduce that

$$\int_{t_0}^{t} W(t')dW(t') = \frac{1}{2}[W(t)^2 - W(t_0)^2 - (t-t_0)]. \tag{2.6.24}$$

2.6.3 Ito's Formula and the Fokker–Planck Equation

The result $dW(t)^2 = dt$ has important implications for how one carries out a change of variables in stochastic calculus. This is most directly established by considering the SDE for an arbitrary function $f(X(t))$ with $X(t)$ evolving according to Eq. (2.6.17):

$$\begin{aligned} df(X(t)) &= f(X(t)+dX(t)) - f(X(t)) \\ &= f'(X(t))dX(t) + \frac{1}{2}f''(X(t))dX(t)^2 + \ldots \\ &= f'(X(t))[a(X,t)dt + b(X,t)dW(t)] + \frac{1}{2}f''(X(t))b(X,t)^2 dW(t)^2, \end{aligned}$$

where all terms of higher order than dt have been dropped. Now using $dW(t)^2 = dt$, we obtain the following SDE for f, which is known as Ito's formula:

$$df(X(t)) = \left[a(X(t),t)f'(X(t)) + \frac{1}{2}b(X,t)^2 f''(X(t))\right]dt + b(X,t)f'(X(t))dW(t). \tag{2.6.25}$$

Hence, changing variables in Ito calculus is not given by ordinary calculus unless f is a constant or a linear function.

We can now use Ito's formula to derive the FP equation for an Ito SDE. First,

$$\begin{aligned} \frac{\langle df(X(t))\rangle}{dt} &= \left\langle a(X(t),t)f'(X(t)) + \frac{1}{2}b(X(t),t)^2 f''(X(t))\right\rangle \\ &= \int \left[a(x,t)f'(x) + \frac{1}{2}b(x,t)^2 f''(x)\right] p(x,t)dx, \\ &= \int f(x)\left[-\frac{\partial}{\partial x}(a(x,t)p(x,t)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(b(x,t)^2 p(x,t))\right]dx, \end{aligned} \tag{2.6.26}$$

after integration by parts, where $p(x,t)$ is the probability density of the stochastic process $X(t)$ under the initial condition $X(t_0) = x_0$. However, we also have

$$\begin{aligned} \frac{\langle df(X(t))\rangle}{dt} &= \left\langle \frac{df(X(t))}{dt}\right\rangle \\ &= \frac{d}{dt}\langle f(X(t))\rangle \\ &= \int f(x)\frac{\partial}{\partial t}p(x,t)dx. \end{aligned} \tag{2.6.27}$$

Comparing Eqs. (2.6.26) and (2.6.27) and using the fact that $f(x)$ is arbitrary, we obtain the Ito version of the FP equation

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}(a(x,t)p(x,t)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(b(x,t)^2 p(x,t)). \tag{2.6.28}$$

2.6.4 Multiplicative Noise and Stratonovich Integrals

It turns out that there is more than one way to define a stochastic difference equation driven by an incremental Wiener process and thus more than one way to obtain an SDE in the continuum limit. This issue only arises in the case of multiplicative noise, that is, when the term multiplying $dW(t)$ depends on the state variable $X(t)$. Recall that in the Ito integral Eq. (2.6.20), it is the value of $b(x,t)$ at the start of the nth time step that multiplies ΔW_n, so that there are no contributions of the form $(\Delta W_n)^2$. An alternative definition of a stochastic integral is the Stratonovich integral

$$\oint_0^T b(X(t),t)dW(t) = \lim_{N\to\infty} \sum_{n=0}^{N-1} b\left(\frac{X_{n+1}+X_n}{2}, t_n\right)\Delta W_n, \tag{2.6.29}$$

where we have used $\oint$ to distinguish it from the Ito integral. Now b depends on the value X_{n+1} at the end of the nth time step, which means there will be an extra term involving $(\Delta W_n)^2$. In order to compare the Ito and Stratonovich integrals, suppose that X_n evolves according to the stochastic difference equation (2.6.18). Thus, in the continuum limit, $X(t)$ is the solution to an Ito SDE. Suppose that we Taylor expand the nth term in the sum defining the Stratonovich integral about the point X_n and set $b_n = b(X_n,t_n)$:

$$b\left(\frac{X_{n+1}+X_n}{2}, t_n\right) = b_n + \frac{\Delta X_n}{2}\frac{\partial b_n}{\partial x} + \frac{1}{2}\left(\frac{\Delta X_n}{2}\right)^2 \frac{\partial^2 b_n}{\partial x^2} + \ldots.$$

Substituting for ΔX_n using Eq. (2.6.18) and dropping terms that are higher order than Δt shows that

$$b\left(\frac{X_{n+1}+X_n}{2}, t_n\right) = b_n + \left(\frac{a_n}{2}\frac{\partial b_n}{\partial x} + \frac{b_n^2}{8}\frac{\partial^2 b_n}{\partial x^2}\right)\Delta t + \left(\frac{b_n}{2}\frac{\partial b_n}{\partial x}\right)\Delta W_n.$$

Applying this result to the sum appearing in the definition of the Stratonovich integral, Eq. (2.6.29), and again dropping higher-order terms in Δt yields the result

$$\sum_{n=0}^{N-1} b\left(\frac{X_{n+1}+X_n}{2}, t_n\right)\Delta W_n = \sum_{n=0}^{N-1} b_n \Delta W_n + \sum_{n=0}^{N-1} \frac{b_n}{2}\frac{\partial b_n}{\partial x}(\Delta W_n)^2.$$

Finally, taking the continuum limit with $dW(t)^2 = dt$, we have

$$\oint_0^T b(X(t),t)dW(t) = \int_0^T b(X(t),t)dW(t) + \frac{1}{2}\int_0^T \frac{\partial b(X(t),t)}{\partial x} b(X(t),t)dt. \tag{2.6.30}$$

Now suppose that $Y(t)$ is a stochastic process evolving according to the Stratonovich SDE

$$dY = a(Y,t)dt + b(Y,t)dW(t). \tag{2.6.31}$$

This means that the integral equation satisfied by $Y(t)$ is based on the Stratonovich integral, that is,

$$Y(t) = y_0 + \int_0^t a(Y(s),s)ds + \oint_0^t b(Y(t),t)dW(t). \tag{2.6.32}$$

Using Eq. (2.6.30), we can rewrite the solution in terms of an Ito integral according to

$$Y(t) = y_0 + \int_0^t \left[a(Y(s),s) + \frac{1}{2}\frac{\partial b(Y(s),s)}{\partial y} b(Y(s),s) \right] ds + \int_0^t b(Y(s),s)dW(s). \tag{2.6.33}$$

The latter is the solution to an equivalent Ito SDE of the form

$$dY = \left[a(Y(t),t) + \frac{b(Y(t),t)}{2}\frac{\partial b(Y(t),t)}{\partial y} \right] dt + b(Y(t),t)dW(t). \tag{2.6.34}$$

Finally, given that we know the FP equation corresponding to an Ito SDE, we can immediately write down the FP equation corresponding to the Stratonovich SDE (2.6.31):

$$\frac{\partial}{\partial t}p(y,t) = -\frac{\partial}{\partial y}(a(y,t)p(y,t)) + \frac{1}{2}\frac{\partial}{\partial y}\left(b(y,t)\frac{\partial}{\partial y}[b(y,t)p(y,t)] \right). \tag{2.6.35}$$

2.6.5 *Ito Integration and Convergence*

So far we have not been specific about the form of convergence used to take the continuum limit of a discrete sum of random variables in order to define a stochastic integral. Following Gardiner [204], we now revisit some results on Ito calculus using the notion of convergence in the mean-square. That is, we define a random variable X to be the limit of a sequence of random variables $\{X_1, X_2, \ldots, X_n\}$ if

$$\lim_{n\to\infty} \left\langle |X - X_n|^2 \right\rangle = 0, \tag{2.6.36}$$

that is, for any $\varepsilon > 0$, there exists an integer $N = N(\varepsilon)$ such that for all $n > N$, $\left\langle |X - X_n|^2 \right\rangle < \varepsilon$. Given this definition of convergence, a stochastic process $X(t)$ is said to be mean-square integrable on the interval $(0,t)$ if there exists a random process $Z(t)$ such that the following limit exists:

$$\lim_{n\to\infty} \left\langle (Z_n - Z(t))^2 \right\rangle = 0, \tag{2.6.37}$$

where

$$Z_n = \Delta t \sum_{j=0}^{n} X(j\Delta t), \quad n\Delta t = t.$$

We then formally write $Z(t) = \int_0^t X(s)ds$.

Suppose that $G(t)$ is a *non-anticipating function*, that is, $G(t)$ is statistically independent of $W(s)-W(t)$ for all $s>t$, where $W(t)$ is a Wiener process. We will show that

$$\int_0^t G(t')[dW(t')]^2 = \int_0^t G(t')dt',$$

in the mean-square sense, that is, Eq. (2.6.37) holds with

$$Z_n = \sum_{j=0}^{n} G_j[\Delta W_j]^2, \quad Z(t) = \int_0^t G(t')dt',$$

where $G_j = G(j\Delta t)$ and $\Delta W_j = W((j+1)\Delta t) - W(j\Delta t)$. Consider

$$I = \lim_{n\to\infty} \left\langle \left[\sum_j G_j(\Delta W_j^2 - \Delta t) \right]^2 \right\rangle$$
$$= \lim_{n\to\infty} \left\langle \sum_j G_j^2(\Delta W_j^2 - \Delta t)^2 + 2\sum_{i>j} G_i G_j(\Delta W_j^2 - \Delta t)(\Delta W_i^2 - \Delta t) \right\rangle.$$

Note that G_j^2 is statistically independent of $(\Delta W_j^2 - \Delta t)^2$ and $G_iG_j(\Delta W_j^2 - \Delta t)$ is statistically independent of $(\Delta W_i^2 - \Delta t)$ for $j<i$. Using the Gaussian nature of ΔW_i, we have

$$\langle \Delta W_i^2 \rangle = \Delta t, \quad \langle (\Delta W_i^2 - \Delta t)^2 \rangle = 2\Delta t^2.$$

Thus we find that

$$I = 2\lim_{n\to\infty} \sum_j G_j^2 \Delta t^2 = 0,$$

assuming that $G(t)$ is bounded. Thus, for Ito integrals $dW(t)^2$ acts like dt.

2.6.6 Simulation of Stochastic Differential Equations

Consider the scalar SDE

$$dX = a(X)dt + b(X)dW(t), \tag{2.6.38}$$

where $W(t)$ is a Wiener process. As with ordinary differential equations, the simplest numerical scheme is to use a direct Euler method. That is, given the solution $X(t)$ at time t, the solution at time $t+\Delta t$ is given by $X(t+\Delta) = X(t) + \Delta X$, where ΔX is determined explicitly by the equation

$$\Delta X = a(X(t))\Delta t + b(X(t))\Delta W, \tag{2.6.39}$$

for a Gaussian random variable ΔW with zero mean and variance equal to Δt. Iterating this process using a random number generator to choose a new value of ΔW at each time step Δt results in an approximation of a sample path of the stochastic process $X(t)$. Repeating the simulation over many independent trials up to a time T then generates a histogram of values of $X(T)$, which can be used to determine an approximation of the probability density for $X(T)$ and to estimate the mean and variance. The direct Euler method is easily extended to multivariate SDEs and those with nonautonomous coefficients a, b.

The accuracy of Euler's method increases with decreasing step size Δt, and the approximate sample path converges in mean-square to the true sample path in the limit $\Delta t \to 0$. For a rigorous discussion of estimating the accuracy of a stochastic numerical algorithm see the book by Kloeden and Platen [338]. Here we give a heuristic definition of the numerical error (see also [204, 298]). Suppose that the time interval $(0, T)$ is divided into N infinitesimal subintervals of size $\tau = T/N$, so that the stochastic process $X(t)$ is sampled at the times τ_n with $n = 0, \ldots, \tau_N$. Let $x_n = X(\tau_n)$ be the exact solution on a given sample path and y_n the corresponding numerical approximation of the solution on the *same* sample path. At the nth time step let $e_n = x_n - y_n$ and define the error at time T to be the root mean-square (RMS) value

$$E(T) = \sqrt{\langle e_N^2 \rangle}. \tag{2.6.40}$$

In the case of the direct Euler method,

$$E(T) \sim \tau^{1/2},$$

and the Euler method is said to be accurate to $\tau^{1/2}$ or that the order of convergence is $\tau^{1/2}$. (In general the order of convergence will depend on how we define the numerical error, that is, the particular measure of convergence. We will restrict ourselves to mean-square convergence.) One practical method for checking the accuracy of a numerical simulation of a given sample path is to repeat the simulation after halving the time step Δt. Suppose that $T = N\Delta t$ and the sample path is generated by the N random increments ΔW_n, $n = 0, \ldots, N-1$. If we then halve the time step, then in order to generate an approximation to the same sample path, it is necessary to produce a set of $2N$ Gaussian random numbers $\Delta \hat{W}_m$, $m = 1, \ldots 2N$, such that

$$\Delta \hat{W}_{2n} + \Delta \hat{W}_{2n+1} = \Delta W_n, \quad n = 0, \ldots, N-1.$$

Given the values ΔW_n, this can be realized by generating N random variables r_n with zero mean and variance $\Delta t/2$, and setting

$$\Delta \hat{W}_{2n} = r_n, \quad \Delta \hat{W}_{2n+1} = \Delta W_n - r_n.$$

One can thus successively halve the time step until errors are within acceptable bounds for the given application. The method can also be used to estimate the rate of convergence.

2.6.7 Milstein's Method

The direct Euler method is a low-order numerical method for SDEs due to the fact that in approximating an SDE one needs to take into account the fact that $dW(t)$ is of order $\sqrt{dt}$. Suppose that we rewrite the scalar SDE (2.6.38) as the integral equation

$$X(t) = X(t_0) + \int_{t_0}^{t} a(X(s))ds + \int_{t_0}^{t} b(X(s))dW(s). \tag{2.6.41}$$

We recover Euler's method by taking $t = t_0 + \Delta t$, with $X(t_0)$ known, and approximating the functions a, b in the interval $s \in (t_0, t_0 + \Delta t)$ according to

$$a(X(s)) \approx a(X(t_0)), \quad b(X(s) \approx b(X(t_0)).$$

In order to obtain a more accurate approximation, we apply Ito's formula (2.6.25) to the functions a and b. For example,

$$\begin{aligned} b(X(s)) = b(X(t_0)) + \int_{t_0}^{s} \left[a(X(s'))b'(X(s')) + \frac{1}{2} b(X(s'))^2 b''(X(s')) \right] ds' \\ + \int_{t_0}^{s} b(X(s'))b'(X(s'))dW(s') \end{aligned} \tag{2.6.42}$$

and similarly for $a(X(s))$. Iterating these equations by successively applying Ito's formula to $a(X(s'))$ and so forth generates an approximation of $a(X(s))$ and $b(X(s))$ in terms of $a(X(t_0)), b(X(t_0))$, higher-order derivatives of $a(x), b(x)$ evaluated at x_0, and a remainder. Substituting such an expansion of $a(X(s))$ and $b(X(s))$ into Eq. (2.6.41) generates a higher-order numerical scheme. The Milstein method is the next higher-order approximation to the stochastic integral Eq. (2.6.41) after Euler's method. It is obtained by substituting

$$a(X(s)) \approx a(X(t_0)), \quad b(X(s)) \approx b(X(t_0)) + b(X(t_0))b'(X(t_0)) \int_{t_0}^{s} dW(s')$$

into Eq. (2.6.41) for $t = t_0 + \Delta t$ and $s \in (t_0, t_0 + \Delta t)$. This yields the following equation for ΔX:

$$\begin{aligned} \Delta X = a(X(t_0)) \int_{t_0}^{t_0+\Delta t} ds + b(X(t_0)) \int_{t_0}^{t_0+\Delta t} dW(s) \\ + \frac{1}{2} b(X(t_0))b'(X(t_0)) \int_{t_0}^{t_0+\Delta t} \int_{t_0}^{s} dW(s)dW(s'). \end{aligned}$$

The double integral can be evaluated using Eq. (2.6.24). That is,

$$\begin{aligned}\int_{t_0}^{t_0+\Delta t}\int_{t_0}^{s} dW(s)dW(s') &= \int_{t_0}^{t_0+\Delta t}[W(s)-W(t_0)]dW(s)\\ &= \int_{t_0}^{t_0+\Delta t} W(s)dW(s) - W(t_0)W(t_0+\Delta t)\\ &= \frac{1}{2}[W(t_0+\Delta t)^2 - W(t_0)^2 - \Delta t] - W(t_0)W(t_0+\Delta t)\\ &= \frac{1}{2}[(\Delta W)^2 - \Delta t].\end{aligned}$$

Hence, we arrive at the Milstein algorithm

$$\begin{aligned}\Delta X = &\left[a(X(t_0)) - \frac{1}{2}b(X(t_0))b'(X(t_0))\right]\Delta t + b(X(t_0)\Delta W\\ &+ \frac{1}{2}b(X(t_0))b'(X(t_0))\Delta W^2.\end{aligned} \tag{2.6.43}$$

It turns out that this algorithm has order Δt accuracy, which improves upon the $\sqrt{\Delta t}$ accuracy of Euler's method.

The complexity of Milstein's method increases when there are multiple noise sources. Consider the multivariate SDE

$$dX_i = a_i(\mathbf{X}(t),t)dt + \sum_{j=1}^{M} B_{ij}(\mathbf{X}(t),t)dW_j(t),$$

where $W_i(t)$ are independent Wiener processes. The Milstein approximation of this equation takes the form

$$\Delta X_i = a_i\Delta t + \sum_{j=1}^{M} B_{ij}\Delta W_j + \sum_{j,k=1}^{M}\left[\sum_{m=1}^{M} B_{mj}\frac{\partial B_{ik}}{\partial X_m}\right]\int_{t_0}^{t_0+\Delta t}\int_{t_0}^{s} dW_j(t')dW_k(s).$$

Unfortunately, only when $j=k$ does the double integral reduce to a simple expression involving the discrete stochastic increments ΔW_j along the lines of the scalar case. However, it can be shown that the symmetrized integral is also reducible according to

$$\int_{t_0}^{t_0+\Delta t}\int_{t_0}^{s}[dW_i(s)dW_j(s') + dW_j(s)dW_i(s')] = \Delta W_i\Delta W_j - \delta_{i,j}(t-t_0).$$

It follows that when the matrix $\mathbf{G}$ satisfies the set of relations (commutative noise)

$$\sum_{m=1}^{M} B_{mj}\frac{\partial B_{ik}}{\partial x_m} = \sum_{m=1}^{M} B_{mk}\frac{\partial B_{ij}}{\partial x_m} \tag{2.6.44}$$

for all i,j,k, the double integral can be symmetrized and Milstein's algorithm becomes

$$\Delta X_i = a_i \Delta t + \sum_{j=1}^{M} B_{ij} \Delta W_j + \frac{1}{2} \sum_{j,k=1}^{M} \left[\sum_{m=1}^{M} B_{mj} \frac{\partial B_{ik}}{\partial X_m} \right] \Delta W_j \Delta W_k$$
$$- \frac{1}{2} \sum_{j,k=1}^{M} \left[\sum_{m=1}^{M} B_{mj} \frac{\partial B_{ij}}{\partial X_m} \right] \Delta t.$$

2.6.8 Runge–Kutter and Implicit Methods

One limitation of the Milstein method is that it requires an evaluation of the first derivative of the function $b(X)$ or its higher-dimensional matrix version. In a similar fashion to deterministic equations, one can use a Runger–Kutta method to eliminate the need to evaluate any derivatives. A first-order method that builds upon the Milstein algorithm has been developed by Platen [338]. It is based on using the approximation

$$b(X)b'(X) \approx \frac{1}{\sqrt{\Delta t}} [b(\hat{X}) - b(X)],$$

where

$$\hat{X} = X + a\Delta t + b\sqrt{\Delta t}.$$

Substituting into Eq. (2.6.43) yields the Milstein–Platen method

$$\Delta X = a\Delta t + b\Delta W + \frac{1}{2\sqrt{\Delta t}} [b(\hat{X}) - b(X)][(\Delta W)^2 - \Delta t]. \tag{2.6.45}$$

Similarly, for a multivariate process, one substitutes into the Milstein method the approximation

$$\sum_{m=1}^{M} B_{mj}(\mathbf{X},t) \frac{\partial B_{ik}(\mathbf{X},t)}{\partial X_m} \approx \frac{1}{\sqrt{\Delta t}} [B_{ij}(\hat{\mathbf{X}}^{(k)}) - B_{ij}(\mathbf{X})]$$

with

$$\hat{X}_i^{(k)} = X_i + a_i \Delta t + B_{ik} \sqrt{\Delta t}.$$

Another issue that numerical methods for solving SDEs share with their deterministic counterparts is instability. This refers to a rapid, exponential increase in numerical error, which can occur spontaneously even though the algorithm appears to be converging to a numerically accurate solution prior to the instability. This feature is a particular problem for "stiff" differential equations, that is, those that have two or more disparate time scales. Often an instability can be fixed by using an *implicit* rather than an explicit method. For example, consider a simple Euler scheme for a single variable,

$$X(t+\Delta t) = X(t) + \Delta X(t) = X(t) + a(X(t),t)\Delta t + b(X(t),t)\Delta W(t).$$

The implicit version is obtained by replacing $X(t)$ with $X(t+\Delta t)$ in the functions a, b:

$$X(t+\Delta t) = X(t) + \Delta X(t) = X(t) + a(X(t+\Delta t),t)\Delta t + b(X(t+\Delta),t)\Delta W(t).$$

This is clearly an implicit equation for $X(t+\Delta t)$, which can be solved numerically using the Newton–Raphson method.

Finally, note that we have focused on the speed and accuracy of numerical methods for generating sample paths of a SDE. Convergence to a sample path is known as *strong convergence*. If one is only interested in properties of the corresponding probability density such as the mean and variance, then these properties are determined by averaging over many sample paths. For a given numerical method, the rate of convergence to the mean or variance tends to differ from the rate of strong convergence and is thus referred to as *weak convergence*.

2.7 Exercises

Problem 2.1 (1D random walk). Consider the probability distribution for a 1D unbiased random walk

$$P_N(r) = \frac{1}{2^N} \frac{N!}{\left(\frac{N+r}{2}\right)!\left(\frac{N-r}{2}\right)!}.$$

Using Stirling's formula

$$\log N! \approx N\log N - N + \frac{1}{2}\ln(2\pi N),$$

derive the Gaussian approximation

$$P_N(r) \sim \frac{1}{\sqrt{2\pi N}} \mathrm{e}^{-r^2/2N}.$$

This result includes a factor of $1/2$ in order to take into account the fact that r is even (odd) when N is even (odd).

Problem 2.2 (Random walk on a lattice). Consider a random walker on a 1D lattice with sites ℓ and displacement distribution $p(\ell)$. The probability $P_n(\ell)$ that the walker is at site $\ell + \ell_0$ after n steps, starting at ℓ_0, satisfies the recurrence relation

$$P_n(\ell) = \sum_{\ell'} p(\ell - \ell') P_{n-1}(\ell').$$

(For a homogeneous random walk, ℓ_0 is arbitrary so we can set $\ell_0 = 0$.) Define the generating function $\Gamma(\ell, z)$ according to

$$\Gamma(\ell, z) = \sum_{n \geq 0} z^n P_n(\ell).$$

(a) Show that the generating function satisfies the equation

$$\Gamma(\ell, z) = \delta_{\ell,0} + z \sum_{\ell'} p(\ell - \ell') \Gamma(\ell', z).$$

(b) Introduce the discrete Fourier transform

$$\hat{\Gamma}(k, z) = \sum_{\ell} \mathrm{e}^{ik\ell} \Gamma(\ell, z)$$

and define the structure function of the walk to be

$$\lambda(k) = \sum_{\ell} \mathrm{e}^{ik\ell} p(\ell).$$

From part (a), show that

$$\hat{\Gamma}(k, z) = 1 + z\lambda(k)\hat{\Gamma}(k, z),$$

so that

$$\hat{\Gamma}(k, z) = \frac{1}{1 - z\lambda(k)}.$$

(c) For a standard RW with $p(\ell) = (\delta_{\ell,1} + \delta_{\ell,-1})/2$, we have $\lambda(k) = \cos(k)$. Using the inverse transform

$$\Gamma(\ell, z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{e}^{-ik\ell} \hat{\Gamma}(k, z) dk,$$

and the result of part (b), evaluate the integral to show that

$$\Gamma(0, z) = (1 - z^2)^{-1/2}.$$

Hint: make the change of variables $t = \tan(k/2)$.

(d) For a general structure function

$$\Gamma(\ell, z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\mathrm{e}^{-ik\ell}}{1 - z\lambda(k)} dk,$$

divergence of the integral is only possible if $\lambda(k_0) = 1$ for some $k = k_0$. If this holds then the integral will be dominated by the region around k_0. Show $\mathrm{e}^{ik_0\ell} = 1$ for all ℓ such that $p(\ell) > 0$ and hence $\lambda(k) = \lambda(k - k_0)$. It follows that the local

behavior of $1-\lambda(k)$ near k_0 is the same as the local behavior of $1-\lambda(k)$ around the origin. Show that for small k and an unbiased RW, $\sum_\ell \ell p(\ell) = 0$, we have

$$\lambda(k) \approx 1 - \frac{\sigma^2 k^2}{2}, \quad \sigma^2 = \sum_\ell \ell^2 p(\ell).$$

Hence, deduce that an unbiased 1D RW is recurrent when the MSD per step is finite.

Problem 2.3 (Chapman-Kolmogorov equation). Consider a stationary stochastic process $X(t)$ with initial condition $X(0) = 0$. Setting $p(x,t|x',t') = p(x-x',t-t')$, the Chapman-Kolmogorov equation takes the form

$$p(x,t) = \int_{-\infty}^{\infty} p(x-y,t-\tau)\, p(y,\tau) dy.$$

(a) Using Fourier transforms, show that the CK equation is satisfied if the characteristic function $G(k,t)$ has the general form $\ln G(k,t) = t g(k)$ for some function $g(k)$.

(b) Suppose that the probability density of the continuous process evolves according to the equation

$$\frac{dp}{dt} = \int_{-\infty}^{\infty} w(x-y)[p(y,t) - p(x,t)] dy,$$

where w is a transition probability per unit time. That is, we have a jump process for a continuous random variable. Using Fourier transforms, obtain the solution

$$p(x,t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left[-ikx + t \int_{-\infty}^{\infty} w(x') \left[\mathrm{e}^{ikx'} - 1 \right] dx' \right] dk.$$

Hence determine $g(k)$ for this process.

Problem 2.4 (Electrodiffusion). The flow of ions through channels in the cell membrane is driven by the combination of concentration gradients and electric fields. If interactions between the ions are ignored, then each ion can be treated as an independent Brownian particle moving under the influence of the electric force $-q\nabla\phi$, where ϕ is the electrical potential and q is the charge on the ion. Multiplying the corresponding FP equation by the number N of ions and using an Einstein relation, we obtain the *Nernst–Planck equation*

$$\frac{\partial c(\mathbf{x},t)}{\partial t} = -\nabla \cdot \mathbf{J}, \quad \mathbf{J}(\mathbf{x},t) = -D\left(\nabla c + \frac{qc}{k_B T} \nabla \phi \right),$$

where c denotes ion concentration. Treating an ion channel as a quasi-one-dimensional domain, this reduces to the 1D equation

$$\frac{\partial c(x,t)}{\partial t} = -\frac{\partial J}{\partial x}, \quad J(x,t) = -D\left(\frac{\partial c}{\partial x} + \frac{qc}{k_B T} \frac{\partial \phi}{\partial x} \right).$$

(a) Suppose that the cell membrane extends from $x = 0$ (inside) to $x = L$ (outside) and denote the extracellular and intracellular ion concentrations by c_e and c_i, respectively. Solve the 1D steady-state Nernst–Planck equation to show that there is zero flux through the membrane if the potential difference $V = \phi_i - \phi_e$ across the membrane is given by the Nernst potential

$$V = \frac{k_B T}{q} \ln\left(\frac{c_e}{c_i}\right).$$

(b) Now suppose that there is a constant nonzero flux J of ions through the channel, and assume for simplicity that the electric field is uniform, that is, $\partial\phi/\partial x = -V/L$. Solving the steady-state Nernst–Planck equation with boundary conditions $c(0) = c_i, c(L) = c_e$, derive the Goldman–Hodgkin–Katz equation for the current density:

$$I \equiv qJ = \frac{D}{L}\frac{q^2 V}{k_B T}\frac{c_i - c_e \exp(-qV/k_B T)}{1 - \exp(-qV/k_B T)}.$$

Check that the Nernst potential is recovered when $J = 0$.

(c) Consider two ion species with opposite charges $q_1 = -q_2 = q$. Applying part (b) to the current for each ion species, derive an expression for the membrane voltage V at which the total ionic current is zero.

Problem 2.5 (Ornstein–Uhlenbeck process). Consider the Ornstein–Uhlenbeck process

$$dX = -kX\,dt + \sqrt{D}\,dW(t),$$

where $W(t)$ is a Wiener process.

(a) Using the solution of the SDE,

$$X(t) = X(0)\mathrm{e}^{-kt} + \sqrt{D}\int_0^t \mathrm{e}^{-k(t-t')}\,dW(t'),$$

show that if the initial condition $X(0)$ is Gaussian distributed with zero mean and variance σ^2 then

$$\langle X(t)\rangle = \langle X(0)\rangle \mathrm{e}^{-kt}, \quad \mathrm{Var}[X(t)] = \sigma^2 \mathrm{e}^{-2kt} + \frac{D}{2k}[1 - \mathrm{e}^{-2kt}].$$

(b) The FP equation for the OU process is

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial [kxp(x,t)]}{\partial x} + \frac{D}{2}\frac{\partial^2 p(x,t)}{\partial x^2}.$$

Taking the fixed (deterministic) initial condition $X(0) = x_0$, the initial condition of the FP equation is

$$p(x,0) = \delta(x - x_0).$$

Introducing the characteristic function (Fourier transform)

$$\Gamma(z,t) = \int_{-\infty}^{\infty} \mathrm{e}^{izx} p(x,t)dx,$$

show that

$$\frac{\partial \Gamma}{\partial t} + kz\frac{\partial \Gamma}{\partial z} = -\frac{D}{2}z^2\Gamma.$$

Use separation of variables to obtain a solution of the form

$$\Gamma(z,t) = \Gamma_0(z\mathrm{e}^{-kt})\mathrm{e}^{-Dz^2/4k}$$

with Γ_0 determined by the initial condition for p. Hence, obtain the result

$$\Gamma(z,t) = \exp\left[-\frac{Dz^2}{4k}(1-\mathrm{e}^{-2kt}) + izx_0\mathrm{e}^{-kt}\right].$$

(c) The probability density $p(x,t)$ can be obtained from $\Gamma(z,t)$ using the inverse Fourier transform

$$p(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-izx}\Gamma(z,t)dz.$$

Substituting for Γ using part (e), show that $p(x,t)$ is a Gaussian with mean and variance

$$\langle X(t)\rangle = x_0\mathrm{e}^{-kt}, \quad \mathrm{Var}[X(t)] = \frac{D}{2k}[1-\mathrm{e}^{-2kt}].$$

(d) Show that the solution to the steady-state FP equation is

$$p_s(x) = (2\pi D/k)^{-1/2}\mathrm{e}^{-kx^2/2D}$$

and that this is consistent with the time-dependent solution in the limit $t \to \infty$.

Problem 2.6 (Additive noise). Solve the SDE

$$dX = -\alpha t^2 X dt + dW(t).$$

(a) Performing the change of variables $Y(t) = X(t)\mathrm{e}^{\alpha t^3/3}$, show that

$$dY(t) \equiv Y(X(t+dt),t+dt) - Y(X(t),t) = \mathrm{e}^{\alpha t^3/3}dW(t).$$

How was the change of variables chosen?

(b) Use part (a) to obtain the solution

$$X(t) = X_0\mathrm{e}^{-\alpha t^3/3} + \int_0^t \mathrm{e}^{-\alpha[t^3-s^3]/3}dW(s).$$

(c) Determine the mean $\langle X(t)\rangle$ and variance $\mathrm{var}[X(t)]$. In particular show that the variance is a dimensionless function of αt^3.

Problem 2.7 (Multivariate Ornstein–Uhlenbeck process). Consider the multivariate SDE

$$dX_i = -\sum_{j=1}^{N} A_{ij} X_j dt + \sum_{j=1}^{N} B_{ij} dW_j(t),$$

where $W_j(t)$ form a set of independent Wiener process:

$$\langle dW_i(t) \rangle = 0, \quad \langle dW_i(t) dW_j(t') \rangle = \delta_{i,j} \delta(t - t').$$

Assume a deterministic initial condition $X_j(0) = \bar{x}_j$.

(a) Show that the solution in vector form is given by

$$\mathbf{X}(t) = \mathrm{e}^{-\mathbf{A}t} \bar{\mathbf{x}} + \int_0^t \mathrm{e}^{-\mathbf{A}(t-t')} \mathbf{B} d\mathbf{W}(t').$$

(b) Introduce the correlation function $\mathbf{C}(t,s) = \langle \mathbf{X}(t), \mathbf{X}^T(s) \rangle$ with components

$$C_{ij}(t,s) = \langle X_i(t), X_j(s) \rangle = \langle [X_i(t) - \langle X_i(t) \rangle][X_j(s) - \langle X_j(s) \rangle] \rangle.$$

Using part (a), show that

$$\mathbf{C}(t,s) = \int_0^{\min(t,s)} \mathrm{e}^{-\mathbf{A}(t-t')} \mathbf{B}\mathbf{B}^T \mathrm{e}^{-\mathbf{A}^T(s-t')} dt'.$$

(c) Introduce the covariance matrix $\Sigma(t) = \mathbf{C}(t,t)$ with components

$$\Sigma_{ij}(t) = \langle [X_i(t) - \langle X_i(t) \rangle][X_j(t) - \langle X_j(t) \rangle] \rangle.$$

Derive the matrix equation

$$\frac{d\Sigma(t)}{dt} = -\mathbf{A}\Sigma(t) - \Sigma(t)\mathbf{A}^T + \mathbf{B}\mathbf{B}^T.$$

Hence, show that if $\mathbf{A}$ has distinct eigenvalues with positive real part, then $\Sigma(t) \to \Sigma_0$ where Σ_0 is the stationary covariance matrix satisfying

$$\mathbf{A}\Sigma_0 + \Sigma_0 \mathbf{A}^T = \mathbf{B}\mathbf{B}^T.$$

Problem 2.8 (1D Fokker–Planck equation with space-dependent variance). Consider the 1D FPE with a space-dependent variance due to multiplicative noise:

$$\frac{\partial P}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial x^2} [D(x) P],$$

with $x \in [-1, 1]$ and reflecting boundary conditions.

(a) Determine the steady-state probability density for general $D(x)$.
(b) Calculate the steady-state probability density when $D(x) = k(a + |x|)$ for $k > 0, a > 1$. What happens when $a \to \infty$?

Problem 2.9 (FPE with multiplicative noise). Write down the (Ito) FP equation for the SDE

$$dX = adt + bxdW,$$

for positive constants a, b and $X \in [0,1]$ with reflecting boundary conditions.

(a) Solve the steady-state FP equation up to a normalization factor.
(b) Calculate the steady-state density for $y = 1/x$ and determine the normalization factor—use the change of random variables formula from Sect. 1.3.
(c) Determine $\langle 1/x \rangle$ as a function of a and b.

Problem 2.10 (Power spectrum). Consider the Langevin equation for a noise-driven, damped harmonic oscillator:

$$m\frac{d^2X}{dt^2} + \gamma\frac{dX}{dt} + kX(t) = 2D\xi(t),$$

where $\xi(t)$ is a Gaussian white noise process with zero mean and covariance

$$\langle \xi(t)\xi(t') \rangle = \delta(t - t').$$

(a) Plot the spectrum of $X(t)$ as a function of the angular frequency ω for $\omega_0 \equiv \sqrt{k/m} = 1$, $2D/m = 1$ and various values of $\beta = \gamma/m$. What happens in the limit $\beta \to 0$? What is the significance of ω_0?
(b) The solution to the Langevin equation can be formally written as

$$X(t) = \int_{-\infty}^{\infty} G(\tau)\xi(t - \tau)d\tau,$$

where $G(\tau)$ is the causal Green's function. Determine the real and imaginary parts of the Fourier transform $\tilde{G}(\omega)$ and plot them as a function of ω for the same parameters as part (a).

Problem 2.11 (FPT for random walks on a lattice). Consider a random walker on a 1D lattice with sites ℓ and displacement distribution $p(\ell)$. The probability $P_n(\ell)$ that the walker is at site ℓ after n steps starting at $\ell_0 = 0$ satisfies the recurrence relation (see Ex. 2.2)

$$P_n(\ell) = \sum_{\ell'} p(\ell - \ell')P_{n-1}(\ell').$$

Let $F_n(\ell)$ denote the probability of arriving at site ℓ *for the first time* on the nth step, given that the walker started at $\ell_0 = 0$.

(a) $P_n(\ell)$ and $F_n(\ell)$ are related according to the recurrence relation

$$P_n(\ell) = \delta_{\ell,0}\delta_{n,0} + \sum_{m=1}^{n} F_m(\ell)P_{n-m}(0), \quad n \geq 0.$$

Explain what this relation means physically.

(b) Show that the corresponding generating functions are related according to

$$\Gamma_F(\ell,z) = \frac{\Gamma_P(\ell,z) - \delta_{\ell,0}}{\Gamma_P(0,z)},$$

where

$$\Gamma_P(\ell,z) = \sum_{n\geq 0} z^n P_n(\ell), \quad \Gamma_F(\ell,z) = \sum_{n\geq 0} z^n F_n(\ell).$$

Hence, use Ex. 2.2c to show that for a standard, unbiased RW

$$\Gamma_F(0,z) = 1 - \sqrt{1-z^2}.$$

(c) Let $R(\ell)$ denote the probability that site ℓ is ever reached by a walker starting at $\ell_0 = 0$:

$$R(\ell) = \sum_{n=1}^{\infty} F_n(\ell) \leq 1.$$

Use part (b) to show that $R(0) = 1$ for an unbiased RW (recurrent rather than transient) while the MFPT $\tau(0)$ to return to the origin is infinite, where

$$\tau(\ell) = \sum_{n=1}^{\infty} nF_n(\ell).$$

Problem 2.12 (FPT for a Brownian particle in a semi-infinite domain). Consider a Brownian particle restricted to a semi-infinite domain $x \in [0,\infty)$ with an absorbing boundary condition at $x = 0$. The FP equation is given by

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2}, \quad 0 < x < \infty,$$

with $p(0,t) = 0$.

(a) Check that the solution of the FP equation for the initial condition $x(0) = x_0$ is

$$p(x,t) = \frac{1}{\sqrt{4\pi Dt}} e^{-(x-x_0)^2/4Dt} - \frac{1}{\sqrt{4\pi Dt}} e^{-(x+x_0)^2/4Dt}.$$

(Such a solution can be derived using the method of images, in which one imagines initially placing a fictitious Brownian particle at the image point $x = -x_0$.)

(b) Show that for large times where $\sqrt{Dt} \gg x_0$, the probability density can be approximated by

$$p(x,t) \approx \frac{1}{\sqrt{4\pi Dt}} \frac{xx_0}{Dt} e^{-(x^2+x_0^2)/4Dt}.$$

(c) Calculate the FPT density $f(x_0,t)$ to reach the origin starting from x_0 by calculating the flux through the origin using part (a):

$$f(x_0,t) = D\frac{\partial p(x,t|x_0,0)}{\partial x}\bigg|_{x=0}.$$

Hence show that when $\sqrt{Dt} \gg x_0$, we have the asymptotic behavior

$$f(x_0,t) \sim \frac{x_0}{t^{3/2}}.$$

Deduce that the MFPT to reach the origin is infinite.

Problem 2.13 (Rotational diffusion). Consider a Brownian particle undergoing diffusion on the circle $\theta \in [-\pi,\pi]$. This could represent the orientation of a bacterium during a single run (see Sect. 2.4). The corresponding FP equation for $p(\theta,t)$

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial \theta^2}, \quad -\pi < \theta < \pi, \quad p(-\pi,t) = p(\pi,t), p'(-\pi,t) = p'(\pi,t),$$

where D is the rotational diffusion coefficient.

(a) Using separation of variables the initial condition $p(\theta,0) = \delta(0)$, show that the solution of the FP equation is

$$p(\theta,t) = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} \mathrm{e}^{in\theta}\mathrm{e}^{-Dn^2t}.$$

(b) If t is sufficiently small then $p(\theta,t)$ is strongly localized around the origin $\theta = 0$. This means that the periodic boundary conditions can be ignored and we can effectively take the range of θ to be $-\infty < \theta < \infty$. That is, performing the rescalings $x = \theta/\varepsilon$ and $\tau = \varepsilon^2 t$, show that $p(\theta,t)$ can be approximated by a Gaussian $p(x,t)$ and deduce the small-time approximation

$$\langle \theta^2 \rangle = 2Dt, \quad t \ll \pi^2/D.$$

(c) What happens in the limit $t \to \infty$?

Problem 2.14 (Diffusion in a sphere). Consider the diffusion equation in a spherical cell of radius R:

$$\frac{\partial u(\mathbf{x},t)}{\partial t} = D\nabla^2 u(\mathbf{x},t), \quad 0 < |\mathbf{x}| < R,$$

with boundary condition $u(|\mathbf{x}| = R,t) = u_1$ and initial condition $u(\mathbf{x},0) = u_0$ with u_0, u_1 constants.

(a) Assume a radially symmetric solution $v(r,t) = u(r,t) - u_1$ so that

$$\frac{\partial v(r,t)}{\partial t} = D\frac{\partial^2 v}{\partial r^2} + \frac{2}{r}D\frac{\partial v}{\partial r}, \quad 0 < r < R,$$

with $v(R,t) = 0$ and $v(r,0) = u_0 - u_1$. Use separation of variables $v(r,t) = V(r)T(t)$ to derive the general solution

$$v(r,r) = \sum_{n=1}^{\infty} c_n \mathrm{e}^{-tDn^2\pi^2/R^2} \frac{1}{r} \sin(n\pi r/R).$$

Hint: in order to solve the boundary value problem for $V(r)$, perform the change of variables $\hat{V}(r) = rV(r)$.

(b) Setting $t = 0$ in the general solution and using $v(r,0) = u_0 - u_1$, determine the coefficients c_n. Hint: you will need to use the identity

$$\int_0^R \sin(n\pi r/R)\sin(m\pi r/R)dr = \frac{R}{2}\delta_{n,m}.$$

(c) Determine an approximation for the concentration $u(0,t)$ at the center of the sphere by taking the limit $r \to 0$, with $r^{-1}\sin(\theta r) \to \theta$. Keeping only the leading order exponential term ($n = 1$), show that the time τ for the center to reach a concentration u^*, $u_1 < u^* < u_0$, is approximately

$$\tau = \frac{R^2}{D\pi^2} \ln \frac{2(u_0 - u_1)}{u^* - u_1}.$$

Problem 2.15 (Computer simulations: Langevin equation). Use the algorithms of Sect. 2.6.6 to solve the following problems in MatLab.

(a) Consider the Ornstein–Uhlenbeck process

$$dX(t) = -\lambda X(t)dt + dW(t), \quad X(0) = x_0,$$

where $W(t)$ is a Wiener process. Use direct Euler to simulate 1,000 trajectories on the time interval $[0,1]$ for $\lambda = 1/2$, $\Delta t = 0.01$ and $x_0 = 1$. Compare the mean and covariance of the trajectories with the theoretical values of Ex. 2.6]

(b) Use Milstein's method to simulate the following SDE on the time interval $[0,1]$

$$dX(t) = -\lambda X(t)dt + \mu X(t)dW(t), \quad X(0) = x_0$$

for $\lambda = 0.1, \mu = 0.1$, and $x_0 = 1$. Compare the cases $\Delta t = 0.1$, $\Delta t = 0.001$, and $\Delta t = 10^{-5}$. Check that the histogram of values at $t = 1$ is similar to the histogram obtained by simulating the exact solution

$$X(t) = x_0 \exp\left[(-\lambda - \mu^2/2)t + \mu W(t)\right].$$

Chapter 3
Stochastic Ion Channels

Ion channels are pore-forming membrane proteins that gate the flow of ions across the cell membrane and the membrane of various intracellular organelles [261, 322]. More than 300 different types of ion channels have been identified across different cell types, which are primarily classified by the nature of their gating and the species of ions passing through the open gates. For example, the opening and closing of *voltage-gated* ion channels depends on the voltage gradient across the plasma membrane, while *ligand-gated* ion channels are open or closed by the binding of ligands to the channel. Both types are particularly prominent components of the nervous system, where voltage-gated ion channels underlie the generation of action potentials and ligand-gated (neurotransmitter activated) ion channels mediate conduction across synapses. They also play a key role in a wide variety of biological processes that involve rapid changes in cells, such as cardiac, skeletal, and smooth muscle contraction, epithelial transport of nutrients and ions, T-cell activation, and pancreatic beta-cell insulin release. Ion channels are thus a frequent target of drug therapies. There also exist *mechanically gated* ion channels, which allow sound, pressure, or movement to cause a change in the excitability of specialized sensory cells and sensory neurons. The stimulation of a mechanoreceptor causes mechanically sensitive ion channels to open and produce a transduction current that changes the membrane potential of the cell—a process known as *mechanotransduction*. An important example of mechanotransduction will be considered in Sect. 5.4, where we describe models of active process in hair cells of the inner ear.

Electrophysiological models of a cell typically assume that the number of ion channels is sufficiently large so that one can determine the average transmembrane currents based on the opening probabilities of individual channels, which is an application of the law of large numbers. However, the resulting deterministic equations cannot account for spontaneous events driven by ion channel fluctuations such as SAPs in a neuron. Another example is the spontaneous release of calcium from the sarcoplasmic reticulum of cardiac cells, which is thought to be related to delayed after depolarizations, which are, in turn, believed to initiate fatal cardiac arrhythmias [401, 414]. Furthermore, the stochastic opening and closing of high-conductance

P.C. Bressloff, *Stochastic Processes in Cell Biology*, Interdisciplinary Applied Mathematics 41, DOI 10.1007/978-3-319-08488-6_3

K^+–Ca^{2+} channels is thought to be responsible for the highly stochastic bursting patterns of isolated pancreatic β cells [586].

In this chapter, we consider stochastic models of ion channel gating, which can be used to study various spontaneous electrophysiological events when coupled to a continuous variable such as membrane voltage or calcium concentration. We begin by describing the chemical kinetics underlying the opening and closing of a single ion channel, distinguishing between voltage-gated and ligand-gated channels (Sect. 3.1). In the latter case we describe the Monod–Wyman–Changeux (MWC) model of cooperative binding. We then construct the master equation for an ensemble of independent two-state ion channels and show how it can be reduced to a FPE using a diffusion approximation (Sect. 3.2). We introduce the notion of a *quasi-potential* for steady-state solutions of the master equation and FP equation. We highlight the fact that the diffusion approximation yields a different quasi-potential from the full master equation, which can lead to exponential errors in the steady-state solution. We then address the important problem of *bistability*, whereby the fraction of open ion channels can exists in two distinct stable states in the deterministic limit. We use the diffusion approximation to analyze noise-induced transitions between these two states in terms of a first passage time problem, and derive an Arrhenius formula for the mean time to escape that depends on the corresponding FP quasi-potential (Sect. 3.3). (The more general theory of noise-induced escape, based on the WKB approximation of solutions to jump Markov processes, will be developed in Chap. 10. The WKB approach yields a much better estimate of quasi-potentials and MFPTs compared to the diffusion approximation.) Two examples of bistability are then considered: spontaneous Ca^{2+} release in oocytes (eggs) and cardiac myocytes (Sect. 3.4) and the generation of SAPs in excitable neurons (Sect. 3.5). In the latter case, the stochastic model takes the form of a stochastic hybrid system in which the piecewise deterministic dynamics of the membrane voltage is coupled to a jump Markov process describing the opening and closing of the ion channels. Finally, in Sect. 3.6 we consider the problem of diffusive escape from a compartment through a stochastic gate, which has been used to model diffusion in the plasma membrane.

3.1 Single Ion Channel Kinetics

3.1.1 Voltage-Gated Ion Channels

The major players in the generation of action potentials are voltage-gated sodium (Na^+) and potassium (K^+) channels. Consider a simple two-state model of a K^+ channel that can exist either in a closed state (C) or an open state (O). Transitions between the two states are governed by a continuous-time discrete Markov process

$$C(\text{closed}) \underset{\beta(v)}{\overset{\alpha(v)}{\rightleftharpoons}} O(\text{open}) \tag{3.1.1}$$

with transition rates $\alpha(v), \beta(v)$ depending on the membrane voltage v. For the moment, we assume that the voltage is fixed; the coupling of ion channel kinetics with voltage dynamics will be considered in Sect. 3.5. In order to understand what such a process means, let $Z(t)$ be a discrete random variable taking values $Z \in \{C, O\}$ and set $P_z(t) = \text{Prob}\,[Z(t) = z]$. From conservation of probability,

$$P_C(t) + P_O(t) = 1.$$

The transition rates then determine the probability of jumping from one state to the other in a small interval Δt:

$$\alpha \Delta t = \text{Prob}\,[Z(t + \Delta t) = O | Z(t) = C], \quad \beta \Delta t = \text{Prob}\,[Z(t + \Delta t) = C | Z(t) = O].$$

It follows that there are two possible ways for the ion channel to enter or leave the closed state:

$$\begin{aligned} P_C(t + \Delta t) &= \text{Prob}[C \to C] P_C(t) + \text{Prob}[O \to C] P_O(t) \\ &= [1 - \alpha \Delta t] P_C(t) + \beta \Delta t P_O(t). \end{aligned}$$

Writing down a similar equation for the open state, dividing by Δt, and taking the limit $\Delta t \to 0$ leads to the pair of equations

$$\frac{dP_C}{dt} = -\alpha P_C + \beta P_O \quad \frac{dP_O}{dt} = \alpha P_C - \beta P_O, \tag{3.1.2}$$

which are equivalent, since $P_0(t) + P_1(t) = 1$.

Equation (3.1.2) has the unique stable steady state

$$P_O = \frac{\alpha}{\alpha + \beta}, \quad P_C = \frac{\beta}{\alpha + \beta}. \tag{3.1.3}$$

Such a steady state has to be consistent with equilibrium statistical mechanics. For a single ion channel maintained at a fixed temperature T, the open and closed probabilities are determined by the Boltzmann–Gibbs distribution (see Sect. 1.4). In particular, we find that $P_O = P_C e^{-\Delta E / k_B T}$, where $\Delta E = E_O - E_C$ is the difference in free energy between the open and closed states. It follows that

$$\frac{\alpha}{\beta} = \mathrm{e}^{-\Delta E / k_B T} \tag{3.1.4}$$

with ΔE a function of membrane voltage. Typically,

$$\Delta E(v) = qv,$$

where the constant q is determined by the displacement of charge when the ion channel changes its conformational state. Now suppose that there are N identical, independent two-state ion channels evolving according to the simple Markov process (3.1.2). In the limit $N \to \infty$ we can reinterpret P_C and P_O as the mean fraction

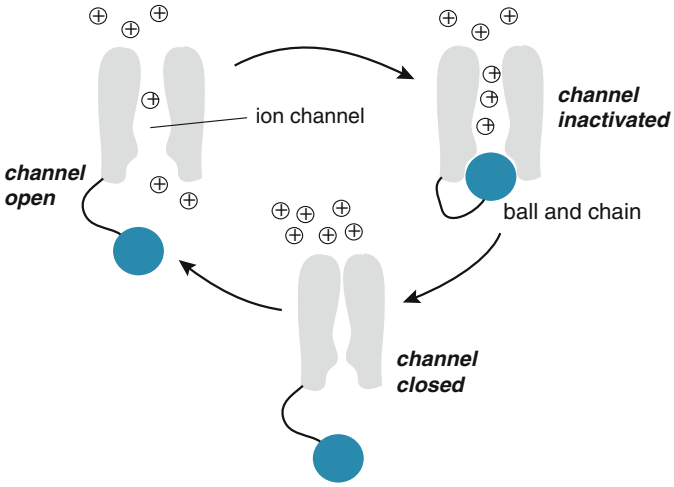


Fig. 3.1: Schematic diagram of the opening/closing of a single sodium ion channel described by a 3-state model. The reduced 2-state model ignores the inactivated state

of closed and open ion channels within the population, and fluctuations can be neglected. After setting $P_O = x$ and $P_C = 1 - x$, we obtain the kinetic equation

$$\frac{dx}{dt} = -\beta x + \alpha(1-x). \tag{3.1.5}$$

The above two-state model is a simplification of more detailed Markov models, in which there can exist inactivated states and multiple subunits [600]. For example, the Na^+ channel inactivates as well as activates (see Fig. 3.1). (An example of an ion channel with multiple internal states is considered in Ex. 3.1.) Moreover, both K^+ and Na^+ channels consist of multiple subunits, each of which can be in an open state, and the channel only conducts when all subunits are open. For example, suppose that a channel consists of two identical, independent subunits, each of which can be open or closed, and that an ionic current can only flow through the channel if both the subunits are open. Let S_j denote the state in which j subunits are open.The transitions between the different states of the ion channel are governed by the reaction scheme

$$S_0 \underset{\beta}{\overset{2\alpha}{\rightleftharpoons}} S_1 \underset{2\beta}{\overset{\alpha}{\rightleftharpoons}} S_2,$$

where α, β are the rates of opening and closing of a single subunit. The factors of two take into account the fact that the state S_0 (S_2) has two closed (open) states either of which can open (close). The corresponding kinetic equations for a large number of identical, independent channels are

$$\frac{dx_0}{dt} = \beta x_1 - 2\alpha x_0, \quad \frac{dx_2}{dt} = \alpha x_1 - 2\beta x_2, \tag{3.1.6}$$

where x_j is the fraction of channels in state j and $x_0+x_1+x_2=1$. The steady-state solution satisfies $x_0^*=\beta x_1^*/(2\alpha)$ and $x_2^*=\alpha x_1^*/(2\beta)$, which implies

$$x_1^*\left(1+\frac{\beta}{2\alpha}+\frac{\alpha}{2\beta}\right)=1.$$

Hence,

$$x_0^*=\frac{\beta^2}{(\alpha+\beta)^2},\quad x_1^*=\frac{2\alpha\beta}{(\alpha+\beta)^2},\quad x_2^*=\frac{\alpha^2}{(\alpha+\beta)^2}.$$

It is straightforward to show that the steady-state solution is stable by linearizing the kinetic equations. (Linear stability analysis for general ODEs is presented in Box 4B. That is, setting $y_j=x_j-x_j^*$ and using $\sum_j y_j=0$, we have

$$\frac{dy_0}{dt}=-\beta(y_0+y_2)-2\alpha y_0,\quad \frac{dy_2}{dt}=-\alpha(y_0+y_2)-2\beta y_2.$$

Introducing the vector $\mathbf{y}=(y_0,y_2)^{\mathrm{tr}}$, this pair of equations can be rewritten in the matrix form

$$\frac{d\mathbf{y}}{dt}=\mathbf{A}\mathbf{y},\quad \mathbf{A}=\begin{pmatrix}-\beta-2\alpha & -\beta\\ -\alpha & -\alpha-2\beta\end{pmatrix}.$$

The linear ODE has solutions of the form $\mathbf{y}=\mathbf{v}e^{\lambda t}$ with $(\lambda,\mathbf{v})$ satisfying the eigenvalue equations

$$\mathbf{A}\mathbf{v}=\lambda\mathbf{v}.$$

This only has nontrivial solutions if $\mathbf{A}-\lambda\mathbf{I}$ is not invertible, where $\mathbf{I}$ is the unit matrix. We thus obtain the characteristic equation

$$0=\det(\mathbf{A}-\lambda\mathbf{I})\equiv(\lambda+\beta+2\alpha)(\lambda+\alpha+2\beta)-\alpha\beta.$$

Rearranging, we have

$$(\lambda+\alpha+\beta)(\lambda+2[\alpha+\beta])=0,$$

and so $\lambda=\lambda_{1,2}$ with $\lambda_1=-(\alpha+\beta)$ and $\lambda_2=-2(\alpha+\beta)$. Since $\lambda_{1,2}<0$, it follows that $\mathbf{y}(t)\to 0$ as $t\to\infty$ and the steady state is stable.

One of the interesting features of models of ion channels with two or more subunits is that the kinetic equations can often be reduced to a lower-dimensional set of equations due to the existence of a stable *invariant manifold*—solutions that start in the manifold cannot leave it and other solutions exponentially converge to the invariant manifold. In the case of the two-subunit model, this can be shown by direct substitution. That is, setting

$$x_0=(1-n)^2,\quad x_1=2n(1-n),\quad x_2=n^2 \tag{3.1.7}$$

and substituting into the kinetic equations (3.1.6) gives

$$-2(1-n)\frac{dn}{dt} = 2\beta n(1-n) - 2\alpha(1-n)^2, \quad 2n\frac{dn}{dt} = 2\alpha n(1-n) - 2\beta n^2,$$

which are both satisfied provided that

$$\frac{dn}{dt} = \alpha(1-n) - \beta n. \tag{3.1.8}$$

Thus, if the initial state can be expressed in terms of the single variable n according to equations (3.1.7), then the solution remains in this one-dimensional space with the dynamics described by the single kinetic equation (3.1.8). Moreover, the stability of the unique steady state implies that the invariant manifold is stable. Since the conducting state of each ion channel corresponds to S_2 (both subunits in the open state), it follows that the fraction of conducting ion channels at time t is $n^2(t)$. Thus the expected conductance of an ion channels is proportional to n^2. Such a result generalizes to more complex ion channel models such as Hodgkin–Huxley (see Ex. 3.2 and Sect. 3.5).

3.1.2 Ligand-Gated Ion Channel

Another very important class of ion channel involves receptor–ligand binding rather than voltage as a gating mechanism. Ligand-gated ion channels are exemplified by neurotransmitter receptors at chemical synapses, where the binding of freely diffusing ligands to a receptor induces a change in conformational state that increases the chance of opening the ion channel. Changes in state of the receptor can be modeled in terms of a set of chemical reactions. Given any chemical reaction (or sequence of reactions) there is a general principle known as the *law of mass action* that determines the form of the kinetic equations describing the evolution of the molecular concentrations. The law states that the rate of an elementary reaction (a reaction that proceeds through only one step) is proportional to the product of the concentrations of the participating molecules. In thermodynamic equilibrium, the rates of the forward and backward reactions must be equal, which allows one to express the ratio of the concentrations of reactants and products in terms of a constant known as the dissociation constant K_d. An expression for K_d can be derived from first principles using the Boltzmann–Gibbs distribution of statistical mechanics (see Sect. 1.4). However, the absolute values of the transition rates (rather than their ratios) cannot be determined from the theory of equilibrium systems. Instead one has to apply some version of Kramers reaction rate theory in order to determine noise-induced transitions between minima of an underlying free energy landscape that represents molecular interactions [204, 253] (see Sect. 3.3). For the moment, we will focus on equilibrium states of the ion channels (see also Chap. 7 of [509]).

We begin by returning to the simple two-state model of an ion channel that was briefly considered in Sect. 1.4. That is, we assume that the ion channel is always open when a single receptor binding site is occupied by a ligand and is always closed when the binding site is unoccupied. The corresponding receptor–ligand reaction is

$$R + L \rightleftharpoons LR,$$

where R denotes an unbound receptor, L is a ligand, and LR is a the receptor–ligand complex. Applying the law of mass action to receptor–ligand binding gives

$$\frac{[LR]}{[R][L]} = \frac{1}{K_d}.$$

Assuming that the total number of receptors is fixed, $[R] + [LR] = [R_{\text{tot}}]$, we have

$$\frac{[LR]}{([R_{\text{tot}}] - [LR])[L]} = \frac{1}{K_d},$$

which on rearranging gives

$$\frac{[LR]}{[R_{\text{tot}}]} = \frac{[L]}{[L] + K_d}. \tag{3.1.9}$$

Comparison with the Boltzmann–Gibbs distribution of equation (1.4.9), with $P_O = [LR]/[R_{\text{tot}}]$ and $[L] = c/c_0$, establishes that the dissociation constant is

$$K_d = \mathrm{e}^{\varepsilon/k_B T},$$

where ε is the binding energy. The fraction of bound receptors increases linearly with $[L]$ at low ligand concentrations but saturates at high concentrations for which $[L] \gg K_d$.

A sharper dependence on $[L]$ can be obtained if there is some form of cooperative binding [518]. The latter refers to situations in which a receptor has multiple binding sites, which can influence each other. An extreme example is when a receptor has n binding sites such that mutual interactions force all of the binding sides to be either simultaneously occupied or simultaneously empty. This can be represented by the reaction scheme

$$R_0 + nL \rightleftharpoons R_n,$$

where R_0 denotes a receptor with empty binding sites and R_n denotes a receptor with all sites filled. The law of mass action shows that at equilibrium

$$\frac{[R_n]}{[L]^n[R_0]} = \frac{1}{K_n},$$

where K_n is an effective dissociation rate. Note that since the forward reaction involves n ligands, one has to include the factor $[L]^n$. Again setting $[R_n] + [R_0] = [R_{\text{tot}}]$ and rearranging gives

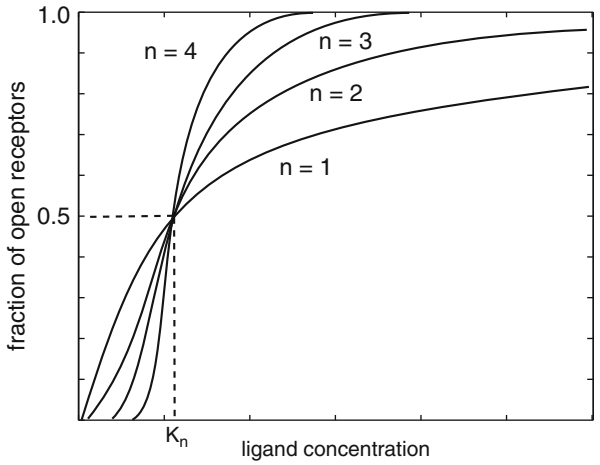


Fig. 3.2: Cooperative binding model. Plot of fraction of open receptors as a function of ligand concentration for various n, where n is the number of binding sites

$$\frac{[LR]}{[R_{\text{tot}}]} = \frac{[L]^n}{[L]^n + K_n}. \tag{3.1.10}$$

The dependence of the fraction of open ion channels as a function of $[L]$ and n is illustrated in Fig. 3.2. Note that ligand-gated ion channels can also exist in more than two conformational states, as considered in Ex. 3.3.

3.1.3 Monod–Wyman–Changeux Model

The above model of receptor–ligand binding is unrealistic in at least two aspects. First, the binding to multiple sites is not all-or-none, that is, a fraction of sites can be occupied at any one time. Second, it is possible for the ion channel to be either open or closed in each binding state—changes in binding state shift the balance between open and closed. A more realistic model of a ligand-gated ion channel with cooperative binding has been developed for the nicotinic acetylcholine receptor, which is found at the neuromuscular junction. It is analogous to the classical MWC model of dimoglobin [451]. The nicotinic receptor has two binding sites for acetylcholine and the equilibrium between the open and closed state of the channel is shifted to the open state by the binding of acetylcholine. A schematic illustration of the different receptor states together with a reaction diagram is shown in Fig. 3.3. In the diagram T_j denotes a closed receptor with j occupied sites and R_j denotes a receptor in the corresponding open state. Also shown is the equilibrium constant (inverse of the dissociation constant) for each of the reversible reactions. In particular, K_T and K_R are the equilibrium constants for binding of an acetylcholine molecule to an individual site of a closed and an open receptor, respectively. The additional factor

of 2 takes into account the fact that there are two unoccupied sites in the forward reaction $T_0 \to T_1$, whereas the additional factor of 1/2 takes into account the fact that there are two occupied sites in the backward reaction $T_2 \to T_1$ (and similarly for R_j). Finally, Y_j is the equilibrium constant associated with the opening and closing of a receptor with j occupied sites.

Applying the law of mass action to each of the reversible reactions leads to the following set of equations for the concentrations:

$$\frac{[R_i]}{[T_i]} = Y_i \tag{3.1.11a}$$

$$\frac{[T_1]}{[L][T_0]} = 2K_T, \quad \frac{[R_1]}{[L][R_0]} = 2K_R \tag{3.1.11b}$$

$$\frac{[T_2]}{[L][T_1]} = K_T/2, \quad \frac{[R_2]}{[L][R_1]} = K_R/2 \tag{3.1.11c}$$

We are interested in the fraction of receptors that are in the open state, which is

$$p_{\text{open}} = \frac{[R_0] + [R_1] + [R_2]}{[R_0] + [R_1] + [R_2] + [T_0] + [T_1] + [T_2]}.$$

Equations (3.1.11b,c) can be used to express $[T_j]$ and $[R_j]$ in terms of $[T_0]$ and $[R_0]$:

$$[T_1] = 2K_T[L][T_0], \quad [T_2] = (K_T[L])^2[T_0], \quad [R_1] = 2K_T[L][R_0], \quad [R_2] = (K_T[L])^2[R_0].$$

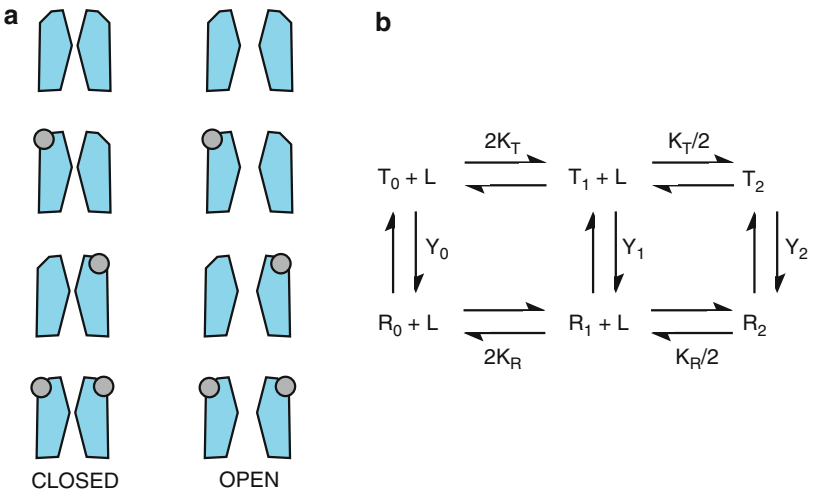


Fig. 3.3: The MWC model of nicotinic acetylcholine receptor with two binding sites. (**a**) Schematic illustration of different conformational states distinguished by the number of occupied binding sites and whether the ion channel is open or closed. (**b**) Reaction diagram

Substituting these results into the formula for p_{open} and using (3.1.11a) gives

$$p_{\text{open}} = \frac{Y_0(1+K_R[L])^2}{Y_0(1+K_R[L])^2+(1+K_T[L])^2}. \tag{3.1.12}$$

We now observe that when $[L]=0$,

$$p_{\text{open}}(0) = \frac{1}{1+1/Y_0},$$

whereas when $[L]$ is large

$$p_{\text{open}}([L]) \approx \frac{1}{1+(K_T/K_R)(1/Y_0)}.$$

It follows that if the open receptor has a higher affinity for binding acetylcholine than the closed receptor ($K_R > K_T$), then $p_{\text{open}}([L]) > p_{\text{open}}(0)$. An interesting feature of MWC type models is that activation of the receptor, as specified by p_{open}, is a sigmoidal function of ligand concentration $[L]$. Thus binding is effectively cooperative even though there are no direct interactions between binding sites. Finally, note that it is straightforward to generalize the MWC model to the case of n binding sites (see Exs. 3.4 and 3.5). Defining the fraction of open receptors according to

$$p_{\text{open}} = \frac{\sum_{j=0}^n [R_j]}{\sum_{j=0}^n [R_j]+\sum_{j=0}^n [T_j]}$$

the law of mass action gives

$$p_{\text{open}} = \frac{Y_0(1+K_R[L])^n}{Y_0(1+K_R[L])^n+(1+K_T[L])^n}. \tag{3.1.13}$$

The MWC model has emerged as a general mechanism for receptor-ligand interactions within a diverse range of applications, including ion channel gating, chemotaxis, and gene regulation (see the review [425] and Sect. 5.3).

3.2 Master Equation for an Ensemble of Ion Channels

Now suppose that there are a finite number N of identical, independent two-state ion channels evolving according to the simple Markov process (3.1.2). In order to take into account fluctuations in the case of finite N, it is necessary to keep track of the probability $P(n,t)$ that there are n open channels at time t, $0 \le n \le N$. (If there are n open channels, then it immediately follows that there are $N-n$ closed channels, so we don't need to keep track of the latter as well.) Consider a time interval $[t, t+\Delta t]$ with Δt sufficiently small so that only one channel has a significant probability of making a $C \to O$ or $O \to C$ transition. There are four possible events that can influence $P(n,t)$ during this interval, two of which involve transitions into

the state of n open ion channels, and two of which involve transitions out of the state. Collecting these terms and taking the limit $\Delta t \to 0$ leads to the *master equation*

$$\frac{d}{dt}P(n,t) = \alpha(N-n+1)P(n-1,t) + \beta(n+1)P(n+1,t) - [\alpha(N-n)+\beta n]P(n,t). \tag{3.2.1}$$

The first term on the right-hand side represents the probability flux that one of $N-(n-1)$ closed channels undergoes the transition $C \to O$, whereas the second term represents the probability flux that one of $n+1$ open channels undergoes the transition $O \to C$. The last two terms represent transitions $n \to n \pm 1$. Define the mean number of open channels at time t by

$$\overline{n}(t) = \sum_{n=0}^{N} nP(n,t).$$

By differentiating both sides of this equation with respect to t and using the master equation (3.2.1) we recover the kinetic equation (3.1.5) with $x = \overline{n}/N$ (see Ex. 3.6).

The two-state ion channel model is an example of a birth–death process described by a master equation of the general form

$$\frac{d}{dt}P(n,t) = \omega_+(n-1)P(n-1,t) + \omega_-(n+1)P(n+1,t) - [\omega_+(n)+\omega_-(n)]P(n,t). \tag{3.2.2}$$

In the case of the simple ion channel model with constant transition rates α, β, we have

$$\omega_+(n) = (N-n)\alpha, \quad \omega_-(n) = n\beta. \tag{3.2.3}$$

However, as we shall see later, more general ion channel models can have transition rates $\omega_\pm(n)$ that are nonlinear functions of n. In the latter case, multiplying both sides of the more general master equation (3.2.2) by n/N and summing over n gives

$$\frac{d\langle n/N\rangle}{dt} = \langle \Omega_+(n/N)\rangle - \langle \Omega_-(n/N)\rangle, \tag{3.2.4}$$

where $\omega_\pm(n) = N\Omega_\pm(n/N)$ and the brackets $\langle \ldots \rangle$ denote a time-dependent ensemble averaging over realizations of the stochastic dynamics, that is,

$$\langle A(n/N)\rangle = \sum_n P(n,t)A(n/N)$$

for any function of state $A(n/N)$. If the transition rates in (3.2.2) are nonlinear functions of n, then there is coupling between different order moments resulting in a moment closure problem. That is, $\langle \Omega_\pm(n/N)\rangle \neq \Omega_\pm(\langle n\rangle/N)$ for finite N. However, in the thermodynamic limit $N \to \infty$, statistical correlations can be ignored so that one can take the mean-field limit

$$\langle \Omega_\pm(n/N)\rangle \to \Omega_\pm(\langle n/N\rangle).$$

This then yields a deterministic equation for the fraction x of open ion channels:

$$\frac{dx}{dt} = \Omega_+(x) - \Omega_-(x). \tag{3.2.5}$$

The steady-state solution $P_s(n)$ of the master equation (3.2.2), assuming it exists, satisfies $J(n) = J(n+1)$ with

$$J(n) = \omega_-(n)P_s(n) - \omega_+(n-1)P_s(n-1).$$

Using the fact that n is a nonnegative integer, that is, $P_s(n) = 0$ for $n < 0$, it follows that $J(n) = 0$ for all n. Hence, by iteration,

$$P_s(n) = P_s(0)\prod_{m=1}^{n}\frac{\omega_+(m-1)}{\omega_-(m)}, \tag{3.2.6}$$

with

$$P_s(0) = \left(1 + \sum_{n=1}^{N}\prod_{m=1}^{n}\frac{\omega_+(m-1)}{\omega_-(m)}\right)^{-1}.$$

In the particular case of the transition rates (3.2.3), we have

$$P_s(n) = P_s(0)\left[\frac{\alpha}{\beta}\right]^n \frac{N!}{n!(N-n)!}. \tag{3.2.7}$$

After calculating $P_s(0)$, we obtain the binomial distribution

$$P_s(n) = \frac{\alpha^n\beta^{N-n}}{(\alpha+\beta)^N}\frac{N!}{n!(N-n)!} = p_0^n(1-p_0)^{N-n}\frac{N!}{n!(N-n)!}, \tag{3.2.8}$$

where $p_0 = \alpha/(\alpha+\beta)$. The mean and variance of the binomial distribution can be obtained using generating functions. That is,

$$\begin{aligned}\Gamma(z) &\equiv \sum_{m=0}^{N} z^m P_s(m)\\ &= \sum_{m=0}^{N}\frac{N!}{n!(N-n)!}(zp_0)^n(1-p_0)^{N-n}\\ &= (zp_0 + 1 - p_0)^N.\end{aligned}$$

It follows that

$$\langle n\rangle = \Gamma'(1) = Np_0(zp_0+1-p_0)^{N-1}\big|_{z=1} = Np_0,$$

and

$$\langle n(n-1)\rangle = \Gamma''(0) = N(N-1)p_0^2(zp_0+1-p_0)^{N-2}\big|_{z=1} = N(N-1)p_0^2.$$

Hence, the mean and variance of the binomial distribution are

$$\langle n\rangle = Np_0, \quad \mathrm{Var}[n] = N(N-1)p_0^2 + Np_0 - N^2p_0 = Np_0(1-p_0). \tag{3.2.9}$$

3.2.1 Diffusion Approximation of a Birth–Death Master Equation

A useful diffusion approximation of the birth–death master equation (3.2.2) for large but finite N can be obtained by carrying out a Kramers–Moyal or system-size expansion to second order in N^{-1} [204, 651], which was originally applied to ion channel models by Fox and Lu [191]. This yields a Fokker–Planck (FP) equation describing the evolution of the probability density of a corresponding continuous stochastic process that is the solution to an SDE. A rigorous analysis of the diffusion approximation has been carried out by Kurtz [360] (see also Chap. 11). First, introduce the rescaled variable $x = n/N$ and transition rates $N\Omega_\pm(x) = \omega_\pm(Nx)$. Equation (3.2.2) can then be rewritten in the form

$$\begin{aligned}\frac{dp(x,t)}{dt} &= N[\Omega_+(x-1/N)p(x-1/N,t) + \Omega_-(x+1/N)p(x+1/N,t)\\ &\quad -(\Omega_+(x)+\Omega_-(x))p(x,t)].\end{aligned}$$

Treating x, $0 \le x \le 1$, as a continuous variable and Taylor expanding terms on the right-hand side to second order in N^{-1} leads to the FP equation

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}[A(x)p(x,t)] + \frac{1}{2N}\frac{\partial^2}{\partial x^2}[B(x)p(x,t)] \tag{3.2.10}$$

with

$$A(x) = \Omega_+(x) - \Omega_-(x), \quad B(x) = \Omega_+(x) + \Omega_-(x). \tag{3.2.11}$$

In the particular case of the two-state ion channel model with transition rates (3.2.3), we have

$$A(x) = \alpha - (\alpha+\beta)x, \quad B(x) = \alpha + (\beta-\alpha)x.$$

The FP equation takes the form of a conservation equation

$$\frac{\partial p}{\partial t} = -\frac{\partial J}{\partial x}, \tag{3.2.12}$$

where $J(x,t)$ is the probability flux,

$$J(x,t) = -\frac{1}{2N}\frac{\partial}{\partial x}[B(x)p(x,t)] + A(x)p(x,t). \tag{3.2.13}$$

The FP equation is supplemented by the no-flux or reflecting boundary conditions at the ends $x = 0, 1$ and a normalization condition,

$$J(0,t) = J(1,t) = 0, \quad \int_0^1 p(x,t)dx = 1. \tag{3.2.14}$$

The FP equation has a unique steady-state solution obtained by setting $J(x,t) = 0$ for all $0 \leq x \leq 1$. The resulting first-order ODE can be solved to give a steady-state probability density of the form

$$P_{\rm FP}(x) = \mathcal{N} \frac{e^{-N\Psi(x)}}{B(x)}, \tag{3.2.15}$$

with the so-called *quasi-potential*

$$\Psi(x) \equiv -2\int^x \frac{A(x')}{B(x')}dx' = -2\int^x \frac{\Omega_+(x') - \Omega_-(x')}{\Omega_+(x') + \Omega_-(x')}dx'. \tag{3.2.16}$$

Here $\mathcal{N}$ is a normalization factor.

Recall from Sect. 2.2 that the solution to the FP equation (3.2.10) determines the probability density function for a corresponding stochastic process $X(t)$, which evolves according to the SDE or Langevin equation [204]

$$dX = A(X)dt + \frac{1}{\sqrt{N}} b(X)dW(t), \tag{3.2.17}$$

with $b(x)^2 = B(x)$. Here $W(t)$ denotes a Wiener process with $dW(t)$ distributed according to a Gaussian process with mean and covariance

$$\langle dW(t)\rangle = 0, \quad \langle dW(t)dW(s)\rangle = \delta(t-s)dtds. \tag{3.2.18}$$

Note that the noise term in (3.2.17) is multiplicative, since it depends on the current state $X(t)$. It is well known that there is an ambiguity in how one integrates multiplicative noise terms, which relates to the issue of Ito versus Stratonovich versions of stochastic calculus [204] (see Sect. 2.6). However, for this particular example, based on the reduction of a master equation, the explicit form of the corresponding FP equation (3.2.10) ensures that the noise should be interpreted in the sense of Ito. In the limit $N \to \infty$, we recover the deterministic equation (3.2.5) with $x(t)$ converging to the unique stable fixed point

$$x^* = \alpha/(\alpha+\beta). \tag{3.2.19}$$

One can thus view the SDE as describing a stochastic path in phase space that involves Gaussian-like fluctuations of order $1/\sqrt{N}$ about the deterministic trajectory. Substituting $X - x^* = Y/\sqrt{N}$ into the SDE (3.2.17) and formally Taylor expanding to lowest order in $1/\sqrt{N}$ yields the so-called *linear noise approximation*

$$dY = -kYdt + b(x^*)dW(t), \tag{3.2.20}$$

with

$$k \equiv -A'(x^*) = \alpha + \beta, \quad b(x^*) = \sqrt{B(x^*)} = \sqrt{\frac{2\alpha\beta}{\alpha+\beta}}.$$

This takes the form of an Ornstein–Uhlenbeck equation [204] [see equation (2.2.16)]. Hence, in the stationary limit $t \to \infty$,

$$\langle Y(t)\rangle \to 0, \quad \langle Y(t)^2\rangle \to \frac{b(x^*)^2}{2k}\left[1 - \mathrm{e}^{-2kt}\right].$$

Since $Y(t) = \sqrt{N}(X(t) - x^*)$, $X(t) = n(t)/N$, and $x^* = p_0$, we recover the results of (3.2.9).

How does the resulting steady-state density given by (3.2.15) with quasi-potential (3.2.16) compare to the steady-state solution of the corresponding master equation (3.2.2) in the large N limit? In order to answer this question, let us consider the particular transition rates (3.2.3). Taking logarithms of both sides of equation (3.2.7) and using Stirling's formula $\log(n!) \approx n\log n - n$ yield a steady-state density similar in form to (3.2.15) but with a different quasi-potential:

$$P(x) = K(x)\mathrm{e}^{-N\Phi(x)}, \tag{3.2.21}$$

with $K(x) = O(1)$ and

$$\begin{aligned}\Phi(x) &= -x\log(\alpha/\beta) + x\log(x) + (1-x)\log(1-x)\\ &= \int^x \ln\frac{\Omega_-(x')}{\Omega_+(x')}dx'.\end{aligned} \tag{3.2.22}$$

Since $\Phi(x) \neq \Psi(x)$, we see that the steady-state probability density under the diffusion approximation can deviate significantly from the effective potential obtained directly from the master equation. However, this discrepancy is not much of an issue for the simple two-state system, since the underlying kinetic equation has a unique fixed point. Indeed, both potentials have the same global minimum at $x = x^*$, $\Phi'(x^*) = \Psi'(x^*) = 0$. Moreover, we find that $\Psi''(x^*) = \Phi''(x^*)$. Since N is large, we can make the Gaussian approximation

$$P(x) \approx p(x^*)\exp\left[-N\Phi(x^*) - N\Phi''(x^*)(x-x^*)^2/2\right],$$

and similarly for $P_{\mathrm{FP}}(x)$. Under this approximation, the mean and variance of the fraction of open channels are given by

$$\frac{\bar{n}}{N} = x^* = \frac{\alpha}{\alpha+\beta}, \quad \frac{\langle(n-\bar{n})^2\rangle}{N^2} = \frac{1}{N\Phi''(x^*)} = \frac{x^*(1-x^*)}{N}, \tag{3.2.23}$$

and we obtain the same results using the Gaussian approximation of $P_{\mathrm{FP}}(x)$. Thus the diffusion approximation accounts well for the Gaussian-like fluctuations around a globally stable fixed point.

On the other hand, it can lead to exponentially large errors when there are multiple stable fixed points. The diffusion approximation still accounts for the effects of fluctuations well within the basin of attraction of a locally stable fixed point, but there is now a small probability that there is a noise-induced transition to the basin of attraction of another fixed point. Since the probability of such a transition is usually of order $\mathrm{e}^{-\tau N}$ with $\tau = O(1)$, except close to the boundary of the basin of attraction, such a contribution cannot be analyzed accurately using standard Fokker–Planck methods [651]. These exponentially small transitions play a crucial role in allowing the system to approach the unique stationary state (if it exists) in the asymptotic limit $t \to \infty$. In other words, for bistable or multistable systems, the limits $t \to \infty$ and $N \to \infty$ do not commute [19, 252, 653]. Later on we will consider two examples of bistability in a population of ion channels: (i) stochastic calcium release in oocytes and cardiac myocytes (Sect. 3.4) and (ii) membrane voltage fluctuations underlying the initiation of SAPs (Sect. 3.5). In the first case, there is bistability in the fraction x of open ion channels arising from the fact that the transition rates in the birth–death process are nonlinear functions of x. This is due to a feedback mechanism involving CICR. On the other hand, bistability in the membrane voltage of a neuron occurs under the assumption that the kinetics of sodium ion channels is relatively fast and potassium kinetics are frozen. In Ex. 3.7, we consider another well-known example of a two-state chemical reaction that exhibits bistability, namely, an autocatalytic reaction [518].

3.3 Population Channel Bistability and Mean Escape Times

If a population of ion channels exhibits bistability in the deterministic limit $N \to \infty$, then a quantity of considerable interest is the mean time for a noise-induced transition from one fixed point to the other when N is finite. We will show how to estimate the escape time using the diffusion approximation of the underlying birth–death master equation (3.2.2) given by the FP equation (3.2.10). First, it is convenient to rewrite the deterministic kinetic equation (3.2.5) for the fraction of open ion channels in the form

$$\frac{dx}{dt} = A(x) = -\frac{dU}{dx}, \tag{3.3.24}$$

where $U(x)$ is a deterministic potential, which is distinct from the quasi-potentials $\Phi(x)$ and $\Psi(x)$. The minima and maxima of the potential $U(x)$ correspond to stable and unstable fixed points of the deterministic dynamics, respectively. Suppose that there are two stable fixed points $x_\pm$ separated by an unstable fixed point x_0 (see Fig. 3.4), and consider the mean time to escape from x_- to x_+; an almost identical calculation holds for the transition $x_+ \to x_-$. Since the system will rapidly approach the state x_+ once it has passed the maximum at x_0, the major contribution to the escape time will be due to the fluctuation-driven transition from x_- to x_0. We can

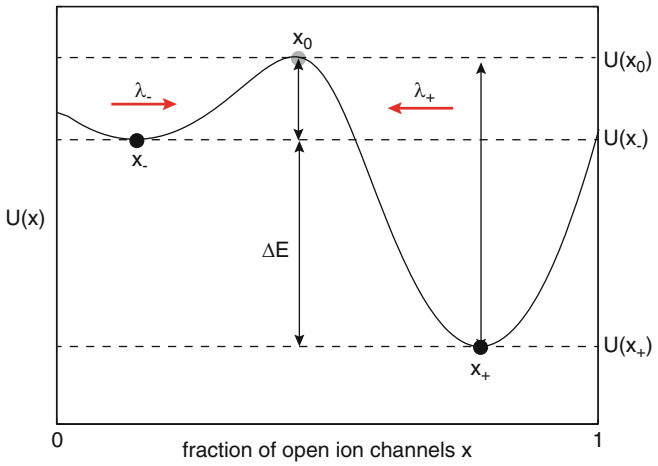


Fig. 3.4: Double-well potential $U(x)$ for a population of two-state ion channels exhibiting bistability in the deterministic limit. Two stable fixed points $x_{\pm}$ are separated by an unstable fixed point x_0

model this process by supplementing the FP equation (3.2.10) with an absorbing boundary condition at $x = x_0$:

$$p(x_0,t) = 0.$$

Let $T(x)$ denote the stochastic time for the particle to exit the right-hand boundary at x_0, given that it starts at location $x \in [0,x_0]$ at time $t = 0$. As a first step, we introduce the survival probability $\mathbb{P}(x,t)$ that the particle has not yet exited at time t:

$$\mathbb{P}(x,t) = \int_0^{x_0} p(x',t|x,0)dx'. \tag{3.3.25}$$

It follows that $\text{Prob}[T(x) \leq t] = 1 - \mathbb{P}(x,t)$ and we can define the FPT density according to

$$f(x,t) = -\frac{\partial \mathbb{P}(x,t)}{\partial t}. \tag{3.3.26}$$

Following along similar lines to the analysis of first passage times in Sect. 2.3, it can be shown that the FPT density satisfies a backward FP equation of the form

$$\frac{\partial \mathbb{P}(x,t)}{\partial t} = A(x)\frac{\partial \mathbb{P}(x,t)}{\partial x} + \frac{B(x)}{2N}\frac{\partial^2}{\partial x^2}\mathbb{P}(x,t). \tag{3.3.27}$$

A quantity of particular interest is the MFPT $\tau(x)$ defined according to

$$\begin{aligned} \tau(x) = \langle T(x)\rangle &\equiv \int_0^{\infty} f(x,t)tdt \\ &= -\int_0^{\infty} t\frac{\partial \mathbb{P}(x,t)}{\partial t}dt = \int_0^{\infty} \mathbb{P}(x,t)dt, \end{aligned} \tag{3.3.28}$$

after integration by parts. Hence, integrating both sides of equation (3.3.27) shows that the MFPT satisfies the ODE

$$A(x)\frac{d\tau(x)}{dx}+\frac{B(x)}{2N}\frac{d^2\tau(x)}{dx^2}=-1. \tag{3.3.29}$$

Equation (3.3.29) is supplemented by reflecting and absorbing boundary conditions for the backward FP equation:

$$\tau'(0)=0, \quad \tau(x_0)=0. \tag{3.3.30}$$

It is straightforward to solve equation (3.3.29) by direct integration [204] (see also Sect. 2.3). First, introducing an integration factor and integrating once gives

$$\mathrm{e}^{-N\Psi(x)}\tau'(x)=-N\int_0^x \frac{\mathrm{e}^{-N\Psi(x')}}{B(x')}dx',$$

where $\Psi(x)$ is the quasi-potential (3.2.16), and we have used the boundary condition $\tau'(0)=0$. Integrating once more with respect to x and using $\tau(x_0)=0$ then gives

$$\tau(x)=N\int_x^{x_0}\mathrm{e}^{N\Psi(x')}dx'\int_0^{x'}\frac{\mathrm{e}^{-N\Psi(x'')}}{B(x'')}dx''. \tag{3.3.31}$$

There is now a standard procedure for approximating this double integral based on Kramers reaction rate theory [204, 253].

For simplicity, let us first consider the FP equation with constant diffusivity

$$\frac{\partial p(x,t)}{\partial t}=-\frac{\partial}{\partial x}[A(x)p(x,t)]+D\frac{\partial^2 p(x,t)}{\partial x^2},$$

where $A(x)=-U'(x)$ and $U(x)$ is given by a double-well potential with minima at $x_\pm$ and a maximum at x_0 with $0<x_-<x_0<x_+$ (see Fig. 3.4). The MFPT to reach a point x beyond x_0, starting from x_-, is

$$\tau_-=\frac{1}{D}\int_{x_-}^{x}\mathrm{e}^{U(x')/D}dx'\int_0^{x'}\mathrm{e}^{-U(x'')/D}dx''.$$

If the central peak of $U(x)$ around x_0 is large and D is small, then $\mathrm{e}^{U(x')/D}$ is sharply peaked around x_0. On the other hand, $\mathrm{e}^{-U(x'')/D}$ is very small near x_0 so $\int_0^{x'}\mathrm{e}^{-U(x'')/D}dx''$ is slowly varying around $x'=x_0$. Hence, $\int_0^{x'}\mathrm{e}^{-U(x'')/D}dx''$ is approximately constant for values of x' such that $\mathrm{e}^{U(x')/D}$ is well above zero. We can thus approximate the MFPT by a product of two independent integrals

$$\tau_-=\frac{1}{D}\left[\int_0^{x_0}\mathrm{e}^{-U(x'')/D}dx''\right]\left[\int_{x_-}^{x}\mathrm{e}^{U(x')/D}dx'\right].$$

The first integral is dominated by a small region around $x = x_-$, whereas the second is dominated by a small region around $x = x_0$. Thus, the integrals are insensitive to the values of the limits and we can take

$$\tau_- = \frac{1}{D}\left[\int_{-\infty}^{\infty} \mathrm{e}^{-U(x'')/D} dx''\right]\left[\int_{-\infty}^{\infty} \mathrm{e}^{U(x')/D} dx'\right].$$

Taylor expanding $U(x'')$ to second order about $x = x_-$ with $U'(x_-) = 0$, we obtain the Gaussian integral

$$\int_{-\infty}^{\infty} \mathrm{e}^{-U(x'')/D} dx'' \approx \int_{-\infty}^{\infty} \mathrm{e}^{-[U(x_-)+U''(x_-)(x''-x_-)^2/2]/D} dx'' = \sqrt{\frac{2\pi D}{U''(x_-)}}\mathrm{e}^{-U(x_-)/D}.$$

Similarly, expanding $U(x')$ to second order about $x = x_0$,

$$\int_{-\infty}^{\infty} \mathrm{e}^{U(x')/D} dx' \approx \int_{-\infty}^{\infty} \mathrm{e}^{[U(x_0)-|U''(x_0)|(x''-x_0)^2/2]/D} dx' = \sqrt{\frac{2\pi D}{|U''(x_0)|}}\mathrm{e}^{U(x_0)/D}.$$

Combining these results, we finally arrive at the classical Arrhenius formula

$$\tau_- \sim \frac{2\pi}{\sqrt{|U''(x_0)|U''(x_-)}}\mathrm{e}^{[U(x_0)-U(x_-)]/D}. \tag{3.3.32}$$

Note that the transition rate $\lambda_- = 1/\tau_-$ (inverse MFPT) from x_- to x_+ varies exponentially with the barrier height $U(x_0) - U(x_-)$, that is, $\lambda_- \sim \mathrm{e}^{-[U(x_0)-U(x_-)]/D}$. Similarly, the transition rate from the active state x_+ to x_- satisfies

$$\lambda_+ \sim \mathrm{e}^{-[U(x_0)-U(x_+)]/D}.$$

It follows that the ratio of the forward and backward transition rates is

$$\frac{\lambda_-}{\lambda_+} = \kappa \mathrm{e}^{[U(x_-)-U(x_+)]/D}, \tag{3.3.33}$$

where κ is a constant that is independent of $U(x_0)$. The exponential dependence on the energy difference $\Delta E = U(x_-) - U(x_+)$ is consistent with equilibrium thermodynamics. For the sake of illustration, suppose that we idealize the above stochastic process as a two-state Markov process involving transitions between two discrete states $O_\pm$ (corresponding to $x = x_\pm$) with rates

$$O_- \underset{\lambda_+}{\overset{\lambda_-}{\rightleftharpoons}} O_+.$$

This is identical in form to the two-state ion channel model evolving according to (3.1.2). Hence, at equilibrium

$$\lambda_- P_- = \lambda_+ P_+,$$

where $P_\pm$ is the steady-state probability of being in state $O_\pm$. Equilibrium thermodynamics requires that $P_\pm$ are given by a Boltzmann–Gibbs distribution (see Sect. 1.4) so that $P_\pm = Z^{-1}\mathrm{e}^{-U_\pm/k_BT}$ where Z is a normalization factor. We deduce that

$$\frac{\lambda_-}{\lambda_+} = \frac{P+}{P_-} = \mathrm{e}^{[U_- - U_+]/k_BT}.$$

Comparison with the previous results shows that $D = k_BT$. Note that this simplified model cannot account for the prefactor κ and its precise form is still a matter of some debate.

Returning to equation (3.3.31), we can adapt the classical Arrhenius formula (3.3.32) with $-\Psi(x)$ an effective potential function and $B(x)/2N$ a state-dependent diffusivity. Note that $\Psi(x)$ has the same stationary points as $U(x)$ so it is also given by a double-well potential. The result is (Ex. 3.8),

$$\tau(x_-) \approx \frac{1}{B(x_-)} \frac{2\pi}{\sqrt{|\Psi''(x_0)|\Psi''(x_-)}} \mathrm{e}^{N[\Psi(x_0)-\Psi(x_-)]}. \tag{3.3.34}$$

Since the argument of the exponential is positive, it follows that the rate of escape is exponentially small for large N. Similarly, the MFPT for the reverse transition $x_+ \to x_-$ is

$$\tau(x_+) \approx \frac{1}{B(x_+)} \frac{2\pi}{\sqrt{|\Psi''(x_0)|\Psi''(x_+)}} \mathrm{e}^{N[\Psi(x_0)-\Psi(x_+)]}. \tag{3.3.35}$$

From our discussion of the diffusion approximation of the steady state, we might expect that a more accurate estimate of the MFPT can be obtained by replacing the quasi-potential $\Psi(x)$ by the quasi-potential $\Phi(x)$ of equation (3.2.22). This is indeed found to be the case, except one also has to modify the prefactor:

$$\tau(x_-) = \frac{2\pi}{\Omega_+(x_-)\sqrt{|\Phi''(x_0)|\Phi''(x_-)}} \mathrm{e}^{N[\Phi(x_0)-\Phi(x_-)]}. \tag{3.3.36}$$

Equation (3.3.36) can be derived using a WKB approximation of the master equation (3.2.2). Such an approach has been used increasingly to analyze escape problems in chemical and biological systems [152, 163, 174, 252, 264, 341]. The details of the calculation for a general birth–death process are presented in Chap. 10.

3.4 Stochastic Models of Ca^{2+} Release

Calcium (Ca^{2+}) is one of the most important and well studied cellular signaling molecules. From a modeling perspective, it attracts a great deal of interest due to the fact that calcium signaling often involves complex spatiotemporal dynamics, including oscillations and waves. For reviews on the modeling of calcium dynamics within cells, see Chap. 7 of Keener and Sneyd [322], Falcke [179], and Chap. 4 of Bressloff [66]. In vertebrates, most of the Ca^{2+} is stored in bones, from where

it can be released by hormonal stimulation to maintain a high extracellular Ca^{2+} concentration (around 1 mM). On the other hand, active ion pumps and exchangers maintain the cytoplasmic Ca^{2+} concentration at relatively low levels (around 10–100 nM). The resulting steep concentration gradient across the plasma membrane means that cells are able to increase their cytoplasmic Ca^{2+} concentration rapidly by opening either voltage-gated or ligand-gated Ca^{2+} ion channels. Here we will consider another major mechanism for controlling intracellular Ca^{2+} based on the action of protein receptors embedded in the surface membrane of intracellular stores. Some of the main features of such receptors are as follows:

1. Cells can regulate their cytoplasmic Ca^{2+} concentration via the intracellular supply of Ca^{2+} from internal stores such as the *endoplasmic reticulum* (ER) and *mitochondria*. *Inositol (1,4,5)-trisphosphate* (IP_3) receptors and *Ryanodine* (Ry) receptors distributed throughout the ER, for example, mediate the release of Ca^{2+} into the cytoplasm, whereas Ca^{2+} ion pumps maintain the relatively high Ca^{2+} concentration within the ER. The Ry receptor plays a critical role in excitation–contraction coupling in skeletal and cardiac muscle cells, but is also found in non-muscle cells such as neurons. One important feature of Ry receptors is that they can undergo CICR, in which elevated cytoplasmic Ca^{2+} activates Ry receptors that release further Ca^{2+}, which then activates other Ry receptors, resulting in a nonlinear regenerative feedback mechanism. The IP_3 receptor is similar in structure to the Ry receptor, but is found predominantly in non-muscle cells and is sensitive to the second messenger IP_3. The binding of an extracellular ligand such as a hormone or a neurotransmitter to a metabotropic receptor results in the activation of a G-protein and the subsequent activation of phospholipase C (PLC). This then cleaves phosphatidylinositol biphosphate (PIP_2) into diacylglycerol (DAG) and IP_3. The water-soluble IP_3 is free to diffuse throughout the cell cytoplasm and bind to IP_3 receptors located on the ER membrane, which then open and release Ca^{2+} from the ER. The opening and closing of an IP_3 receptor is also modulated by the concentration of cytoplasmic Ca^{2+}, so it too can undergo CICR.

2. Another mechanism for controlling cytoplasmic Ca^{2+} is through buffering (binding) to large proteins. It is estimated that at least 99 % of the total cytoplasmic Ca^{2+} is bound to buffers. A summary of the basic extracellular and intracellular mechanisms for controlling cytoplasmic Ca^{2+} is shown in Fig. 3.5.

3. One of the most dramatic consequences of CICR is the propagation of intracellular Ca^{2+} waves mediated primarily by the opening of IP_3 receptors. These waves were first observed in nonneuronal cells such as *Xenopus laevis* oocytes [376, 499], where the resulting changes in Ca^{2+} concentration across the whole cell provided a developmental signal.

4. Many cell types exhibit spontaneous localized Ca^{2+} release events known as *sparks* or *puffs* [103, 104]. The fluorescent imaging of Ca^{2+} puffs and sparks has established that Ca^{2+} release is a stochastic process that occurs at spatially discrete sites consisting of clusters of IP_3Rs and RyRs, respectively. Ca^{2+} puffs

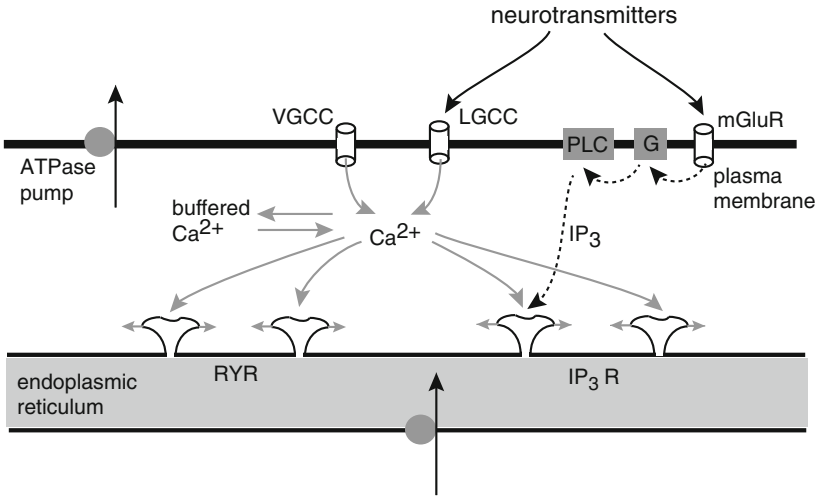


Fig. 3.5: Calcium signaling. The entry of Ca^{2+} from outside the cell is mediated by voltage-gated (VGCC) and ligand-gated (LGCC) calcium channels. Stimulation of metabotropic glutamate receptors (mGluRs) produces IP_3 second messengers that bind to IP_3 receptors (IP_3Rs), which subsequently release Ca^{2+} from the endoplasmic reticulum. Both IP_3Rs and Ryanodine receptors (RyRs) are sensitive to Ca^{2+}, resulting in calcium-induced calcium release (CICR). The latter can sometimes result in the propagation of a Ca^{2+} wave along the dendrites

are found in *Xenopus laevis* oocytes and have an amplitude ranging from around 50–600 nM, a spatial spread of approximately 6 μm and a typical duration of 1 s [498, 499, 692]. For sufficiently high levels of IP_3 concentration, the amplification of Ca^{2+} puffs by CICR can lead to the formation of Ca^{2+} waves [498, 499, 692]. Calcium sparks, which are thought to be the building blocks of the large regenerative Ca^{2+} signal that controls contraction in cardiac and skeletal muscle cells, arise from the opening of clusters of RyRs by local CICR. The frequency of calcium spark events is sensitive to changes in membrane potential, although they rarely induce calcium waves due to shorter duration and less spatial spread.

3.4.1 Stochastic Model of Ca^{2+} Puffs in a Cluster of IP_3Rs

Stochastic models of Ca^{2+} puffs typically treat a cluster of IP_3Rs as a set of N channels that open and close independently, but are indirectly coupled by the common cytoplasmic Ca^{2+} concentration [178, 180, 590, 618]. Models differ in the level of detail regarding individual receptors. The first deterministic kinetic model of Ca^{2+}-gated IP_3Rs was proposed by De Young and Keizer [696], in their study of agonist-induced Ca^{2+} oscillations. This model assumes that the IP_3 receptor consists of

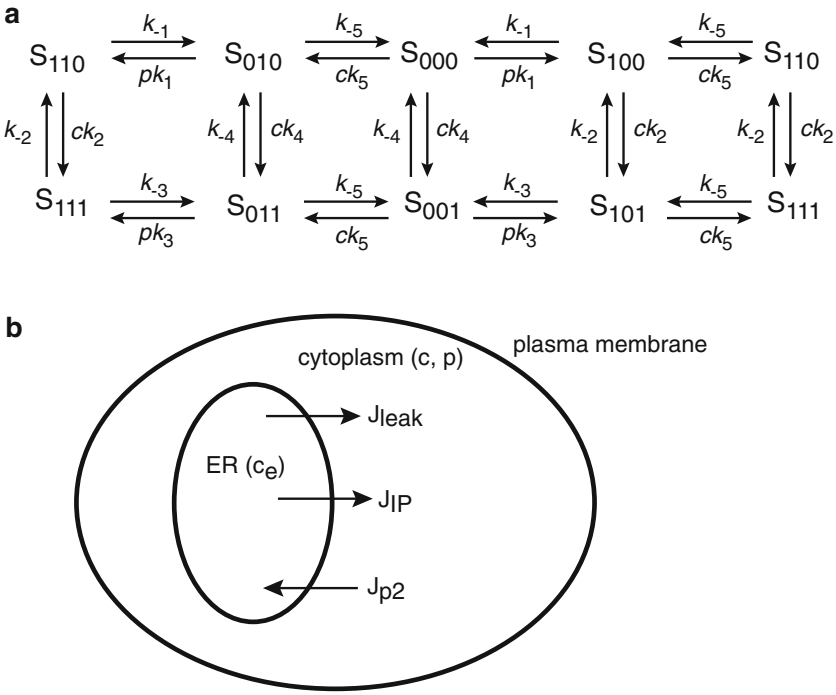


Fig. 3.6: IP_3 receptor model. (**a**) Reaction diagram of the De Young–Keizer model [696] of an IP_3 receptor subunit. The state of the subunit is denoted by S_{ijk}, where $i = 0, 1$ denotes whether the IP_3 binding site is unoccupied ($i = 0$) or occupied ($i = 1$), and j, k denote the corresponding status of the activating and inactivating Ca^{2+} binding sites, respectively. Although there are 24 separate single-step reactions, the model only has ten independent rate constants. This is a consequence of equilibrium thermodynamics and two additional constraints: (i) the rate constants are taken to be independent of whether or not the Ca^{2+} activating binding site is occupied and (ii) the kinetics of Ca^{2+} activation are assumed to be independent of IP_3 binding and Ca^{2+} inactivation. (**b**) Schematic diagram of fluxes in the Li–Rinzel model [388]. Here c and p denote the concentration of Ca^{2+} and IP_3 in the cytoplasm, and c_e is the concentration in the endoplasmic reticulum (ER). Both p and c_e are held fixed

three equivalent receptor subunits, all of which have to be in a conducting state in order to generate a Ca^{2+} flux. Each subunit is taken to have an IP_3 binding site, an activating Ca^{2+} binding site, and an inactivating Ca^{2+} binding site; the conducting state corresponds to the state in which all subunits have the first two binding sites occupied but the third unoccupied (see Fig. 3.6a). Although the De Young–Keizer model is simple to describe, it involves a relatively large number of variables that have to be coupled to the Ca^{2+} and IP_3 concentrations. A simplified version of the model was subsequently developed by Li and Rinzel [388]. They exploited the fact that the binding of IP_3 and activating Ca^{2+} are fast relative to inactivating Ca^{2+} and used a quasi-steady-state (QSS) argument to reduce the eight-state subunit model to a model that simply keeps track of whether or not the inactivating Ca^{2+} binding site

of a subunit is occupied. There is then a single gating variable

$$h = x_{000} + x_{010} + x_{100} + x_{110},$$

where x_{ijk} denotes the fraction of subunits in state S_{ijk} (see Fig. 3.6a). Thus, h^3 is the fraction of receptors in a cluster not inactivated by Ca^{2+}. One finds that h evolves according to an equation of the form (Ex. 3.9)

$$\frac{dh}{dt} = \alpha_h(p)(1-h) - \beta_h(p)ch, \tag{3.4.1a}$$

where c is the cytoplasmic Ca^{2+} concentration and p is the IP_3 concentration (which is assumed fixed). It is assumed that there are three fluxes contributing to the change in Ca^{2+} concentration:

$$\frac{dc}{dt} = J_{\rm IP} + J_{\rm leak} - J_{\rm p2}, \tag{3.4.1b}$$

where $J_{\rm IP}$ is the flux through the cluster of IP_3 receptors, $J_{\rm leak}$ is a leakage flux from the ER to the cytoplasm, and $J_{\rm p2}$ is the flux pumped back into the ER (see Fig. 3.6b). The expressions for the various fluxes are

$$J_{\rm IP} = f(c,p)^3 h^3 [c_e - c], \quad J_{\rm leak} = \nu_0 [c_e - c], \quad J_{\rm P2} = \frac{\nu_1 c^2}{k_3^2 + c^2}, \tag{3.4.2}$$

where c_e is the fixed Ca^{2+} concentration in the ER, ν_0, ν_1 are constants, and

$$f(c,p) = \left(\frac{p}{p+K_1}\right) \cdot \left(\frac{c}{c+K_5}\right), \quad K_j = k_{-j}/k_j. \tag{3.4.3}$$

The function $f(c,p)$ can be derived from the QSS reduction of the De Young–Keizer model, which shows that the fraction of open subunits is $x_{110} = f(c,p)h$ (Ex. 3.9). The cubic terms reflect the existence of three subunits. Parameter values of the model can be found in [388]. Note that the simplified model resembles the Hodgkin–Huxley model of a neuron, which will be introduced in Sect. 3.5 [see equation (3.5.6)] after replacing Ca^{2+} concentration c by membrane voltage v and c_e by a reversal potential.

We now describe a stochastic version of the Li–Rinzel model for a cluster of IP_3Rs due to Shuai and Jung [590]. For stochastic versions of the full De Young–Keizer model, see, for example, [178, 180, 243, 618]. The deterministic equations (3.4.1) describe the mean behavior of a large cluster of Ca^{2+} channels, just as the Hodgkin–Huxley equations for membrane voltage apply to a large number of voltage-gated ion channels. If the number of channels is relatively small, then it is necessary to take into account thermally driven fluctuations in the opening and closing of individual channels. In the case of the Li–Rinzel model, one only needs to

consider the state of the Ca^{2+}-inactivating binding site of each subunit. The latter is modeled as the two-state Markov process

$$A^* \underset{\beta_h(c)}{\overset{\alpha_h}{\rightleftharpoons}} A, \tag{3.4.4}$$

where A (A^*) denotes the unbound (bound) state. Suppose that there are N independent IP_3Rs, each with three independent subunits labeled $i = 1,2,3$ that are described by the above two-state Markov process. Let $N_i(t)$ ($i = 1,2,3$) denote the number of receptors at time t that have the ith subunit in state A. Under the adiabatic assumption that the Ca^{2+} concentration c evolves much more slowly than the state transitions of the channels, we can write down a master equation for the probability $P(n_i,t) = \text{Prob}[N_i(t) = n_i | N_i(0) = n_0]$ according to

$$\begin{aligned}\frac{dP(n_i,t)}{dt} &= (N-n_i+1)\alpha_h P(n_i-1,t) + (n_i+1)c\beta_h P(n_i+1,t) \\ &\quad - (n_i c\beta_h + (N-n_i)\alpha_h)P(n_i,t), \quad i = 1,2,3.\end{aligned} \tag{3.4.5}$$

As with voltage-gated ion channels (see Sect. 3.5), we have a stochastic hybrid system, since the $A \to A^*$ transition rate depends on the Ca^{2+} concentration $c(t)$, which evolves according to a piecewise deterministic equation of the form (3.4.1b). The latter, in turn, couples to the discrete stochastic variables $N_i(t)$ through the flux

$$J_{\text{IP}} = f(c(t),p)[c_e - c(t)] \prod_{i=1}^{3} \frac{N_i(t)}{N}. \tag{3.4.6}$$

Finally, for large N, one can obtain a further simplification by carrying out a Kramers–Moyal expansion of the master equation (3.4.5) along identical lines to Sect. 3.2. This yields the following SDE for $H_i(t) = N_i(t)/N$ with H_i treated as a continuous stochastic variable:

$$dH_i = \alpha_h(1-H_i) - c\beta_h H_i + \frac{1}{\sqrt{N}} b(H_i) dW_i, \tag{3.4.7}$$

where

$$b(H_i) = \sqrt{\alpha_h(1-H_i) + c\beta_h H_i},$$

and $W_i(t)$ is an independent Wiener process with

$$\langle dW_i(t) \rangle = 0, \quad \langle dW_i(t) dW_j(t') \rangle = \delta(t-t') dt\, dt' \delta_{i,j}.$$

Shuai and Jung [590] simulated the stochastic Li–Rinzel model in order to investigate the effects of noise on Ca^{2+} oscillations in a space-clamped model. They assumed that the deterministic system (3.4.1) was monostable at low and high IP_3 concentrations and exhibited limit cycle oscillations (occurring via a Hopf bifurcation, see Box 4B) at intermediate concentrations. They showed that noise can enlarge the range of IP_3 concentrations over which oscillations occur—an effect known as *coherence resonance*. They also found a broad distribution of puff amplitudes,

lifetimes, and interpuff intervals. In particular, at low IP_3 concentrations, the amplitude distribution is a monotonically decaying function, whereas at higher concentrations it is unimodal. This suggests that Ca^{2+} puffs become more significant as IP_3 concentration is increased and hence could impact the spontaneous generation of Ca^{2+} waves. This issue was investigated numerically by Falcke [178] using a stochastic version of the De Young–Keizer model that was incorporated into a reaction–diffusion model of spatially distributed channel clusters. He showed that there is indeed a transition from Ca^{2+} puffs to waves as the IP_3 concentration is increased. At low concentrations, only puffs occur, since there is not enough Ca^{2+} released to stimulate neighboring clusters, which means that the response is purely local. However, as IP_3 concentration increases, global Ca^{2+} waves can emerge from local nucleation sites of high Ca^{2+} concentration. At intermediate levels of IP_3, global events are rare and waves only progress a short distance before dying out. On the other hand, for higher IP_3 concentrations, global waves occur regularly with a well-defined period. Again this oscillatory-like behavior can occur in parameter regimes for which the deterministic model is non-oscillatory.

3.4.2 Stochastic Model of Ca^{2+} Sparks in Cardiac Myocytes

We now turn to a stochastic model of Ca^{2+} sparks in cardiac myocytes [263], which includes details of the geometry of Ca^{2+} release units, in particular, the narrow junctional gap known as the diadic space that separates the sarcoplasmic reticulum (SR) from the plasma membrane (see Fig. 3.7). (In smooth muscle cells the SER is referred to as the sarcoplasmic reticulum.) In a typical myocyte, there could be up to 10,000 Ca^{2+} release units, each one containing a cluster of around $N = 50$ RyRs on the surface of the SR. The cluster of RyRs is apposed to L-type Ca^{2+} channels located on so-called t-tubules, which are invaginations of the plasma membrane into the myocyte. (The Ca^{2+} channels are not involved in the spontaneous generation of Ca^{2+} sparks so are ignored in the model.) The diadic space separating the SR from the t-tubules is a region of the mytoplasm (intracellular fluid of myocytes), which is approximately cylindrical in shape with height 10 nm and radius 100nm. Since the diadic space is a small enclosed volume, it supports an elevation in Ca^{2+} concentration relative to the bulk mytoplasm following the release of Ca^{2+} from an RyR. Such a local elevation plays a crucial role in the CICR that results in a Ca^{2+} spark. The SR in a neighborhood of the RyRs is known as the junctional SR (JSR), which may have a different Ca^{2+} concentration from the bulk or network SR (NSR).

We present the model in nondimensional form; details of model approximations and estimates of experimentally based model parameters can be found in [263]. First, the diadic space is modeled as a single compartment with Ca^{2+} concentration c satisfying the current conservation equation

$$\tau_D \frac{dc}{dt} = J_{\mathrm{RyR}} - J_{\mathrm{D}}. \tag{3.4.8}$$

Here τ_D is a time constant, J_{RyR} is the total Ca^{2+} current through the RyRs, and J_{D} is the diffusive current from the diadic space to the bulk mytoplasm. The latter is modeled as the Fickian current

$$J_{\mathrm{D}} = c - c_m, \tag{3.4.9}$$

where c_m is the bulk mytoplasm Ca^{2+} concentration. The total current through the RyRs is taken to be proportional to the number n of open RyRs times the Ca^{2+} concentration c_{sr} in the JSR:

$$J_{\mathrm{RyR}} = c_{\mathrm{sr}} x, \quad x = \frac{n}{N}, \tag{3.4.10}$$

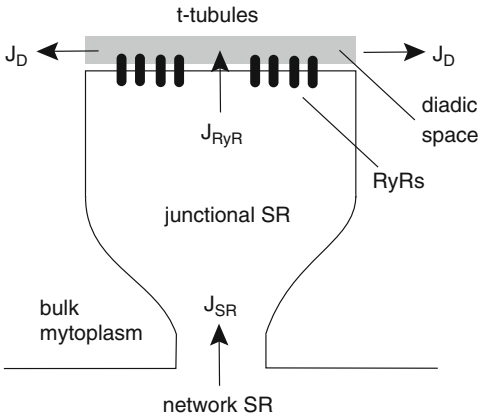


Fig. 3.7: Schematic diagram of a Ca^{2+} release unit in a cardiac myocyte. It is divided into four compartments: the network SR, the junctional SR, the diadic space, and the bulk mytoplasm. See text for details

with N the total number of RyRs in the cluster. Each RyR has Ca^{2+} binding sites, which can be activating or deactivating. When an RyR is in an activated state it is promoted to a mode where it continuously opens and closes according to a Markov process, with a mean open time of 1 ms [699]. The opening of an RyR channel results in an extra Ca^{2+} current flowing into the diadic space, which increases the rate at which Ca^{2+} binds to the other RyRs via CICR, thus creating a positive feedback loop. This feedback loop provides a mechanism for bistability. Note that the RyRs also contain inactivating Ca^{2+} binding sites, but these do not play a role in initiating a Ca^{2+} spark so are not included in the model. For simplicity, the RyRs are modeled using a two-state Markov process involving a single closed state and a single open state (see also [324]):

$$C(\text{closed}) \underset{k_-}{\overset{k_+(c)}{\rightleftharpoons}} O(\text{open}), \tag{3.4.11}$$

with transition rates

$$k_+(c) = \frac{1}{k\tau_o}\frac{c^\alpha}{c^\alpha + 1}, \quad k_- = \frac{1}{\tau_o}. \tag{3.4.12}$$

Here τ_o is the mean open time of a RyR, α is the number of Ca^{2+} ions that are needed to open a RyR, and k determines the proportion of time the RyRs are open. It is assumed that the RyRs are gated independently of each other. They are, however, indirectly coupled via the Ca^{2+} concentration in the diadic space. The time constant τ_D of diffusive flux from the diadic space is several orders of magnitude smaller than the mean open time τ_o of a RyR, that is, $\tau_D \sim 3\mu$s whereas $\tau_o \sim 1$ ms. Therefore, the Ca^{2+} concentration in the diadic space can be taken to be in quasi-equilibrium, $\tau_D \to 0$, so that

$$c = c_m + c_{sr}x. \tag{3.4.13}$$

It follows that the transition rate can be reexpressed as a function of the fraction of open channels and the Ca^{2+} concentration in the SR, $k_+ = k_+(c_m + c_{sr}x)$.

Now consider N independent RyRs within a Ca^{2+} release unit, each described by the above two-state Markov process. Let $N(t)$ be the number of open channels at time t and set $P(n,t) = \text{Prob}[N(t) = n|N(0) = n_0]$. The distribution $P(n,t)$ evolves according to the birth–death master equation (3.2.2) with transition rates $\omega_\pm(n/N)$: $n \to n \pm 1$:

$$\omega_+(x) = N(1-x)\frac{(c_m + c_{sr}x)^\alpha}{k((c_m + c_{sr}x)^\alpha + 1)}, \quad \omega_-(x) = Nx. \tag{3.4.14}$$

For the moment, it is assumed that c_m and c_{sr} are fixed so that ω_+ can be treated as a function of x alone. (Later the dynamics of c_{sr} following initiation of a Ca^{2+} spark will also be taken into account.) The units of time are fixed by setting $\tau_o = 1$. In the deterministic limit $N \to \infty$, we obtain the kinetic equation (3.2.5), which takes the explicit form

$$\frac{dx}{dt} = \Omega_+(x) - \Omega_-(x) = (1-x)\frac{(c_m + c_{sr}x)^\alpha}{k((c_m + c_{sr}x)^\alpha + 1)} - x. \tag{3.4.15}$$

It can be shown that, for physiologically reasonable parameter values, this equation exhibits bistability [263], that is, there exists a pair of stable fixed points $x_\pm$ separated by an unstable fixed point x_0. The fixed point $x_- \approx 0$ represents a quiescent state, whereas the other fixed point x_+ represents a Ca^{2+} spark in which a significant fraction of RyRs are in the active mode and can be interpreted as a burst phase.

Noise-induced transitions from x_- to x_+ determine the distribution of inter-spark intervals, just as noise-induced transitions from x_+ to x_- determine the distribution of spark lifetimes. Hence, estimating the mean time for the occurrence of a spark event reduces to the problem of calculating the MFPT to reach x_+, starting from a neighborhood of x_-, by crossing x_0. This calculation was carried out in Sect. 3.3 for a general birth–death master equation using a diffusion approximation and can be made more accurate using WKB methods (see Chap. 10). The latter approach yields equation (3.3.36) for the mean time τ_i to initiate a Ca^{2+} spark starting from the

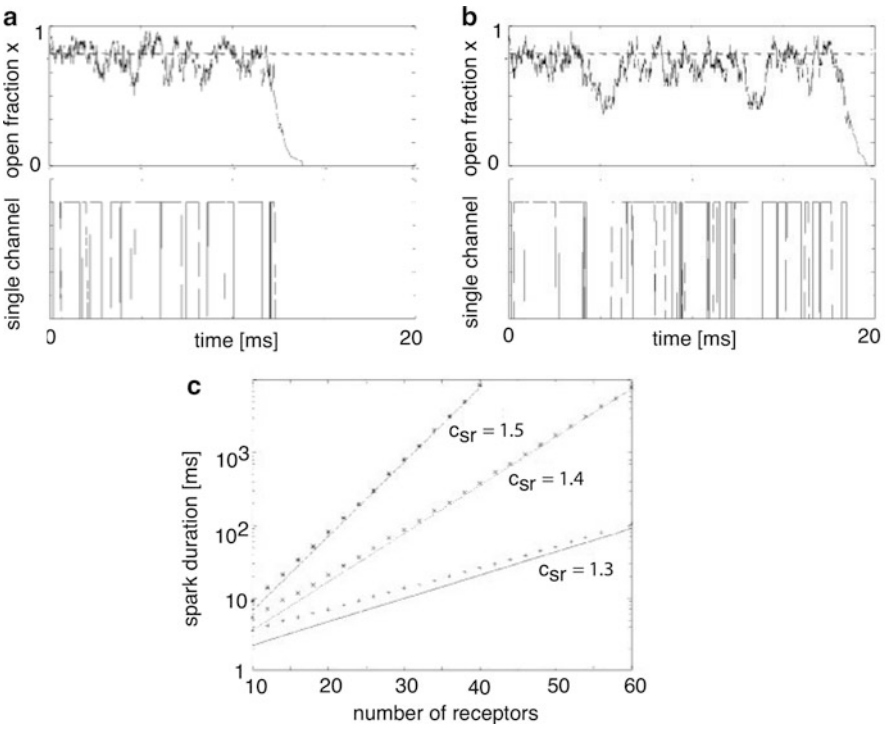


Fig. 3.8: Stochastic model of calcium sparks. (**a,b**) Monte Carlo simulations of the proportion x of RyRs in the open state and the transitions of one of the RyRs in the cluster. (**c**) The mean spark time as a function of the number N of receptors in the cluster. The line is the asymptotic calculation and the points are from a Monte Carlo simulation. As the number of receptors in the cluster increases, the spark time increases and the error in asymptotic calculation decreases. The mean spark length also increases rapidly with SR concentration $c_{\rm sr}$. The model parameters are $\alpha = 4$, $c_m = 0$, and $k = 0.2$. (Adapted from Hinch [263])

quiescent state x_- after substituting for the transition rates using equation (3.4.14). Similarly, the mean duration τ_f of a spark, which corresponds to the mean time to transition back from x_+ to x_-, is given by equation (3.3.36) under the mapping $x_- \to x_+$. It turns out that in the case of Ca^{2+} release, the quiescent state x_- is in an $O(1/N)$ neighborhood of the boundary $x = 0$, so that the prefactor of the MFPT has to be modified accordingly (see [263, 264] for details). Nevertheless, the leading order exponential is unchanged. Hinch compared the theoretical prediction with Monte Carlo simulations of the full system for various $c_{\rm st}$, and the results are shown in Fig. 3.8. One can see that the spark time increases with the number of receptors in the cluster and the mean spark length increases rapidly with $c_{\rm sr}$.

Irrespective of the particular method used to solve the FPT problem, it was assumed above that the concentration c_{sr} in the JSR is held fixed. This is a reasonable

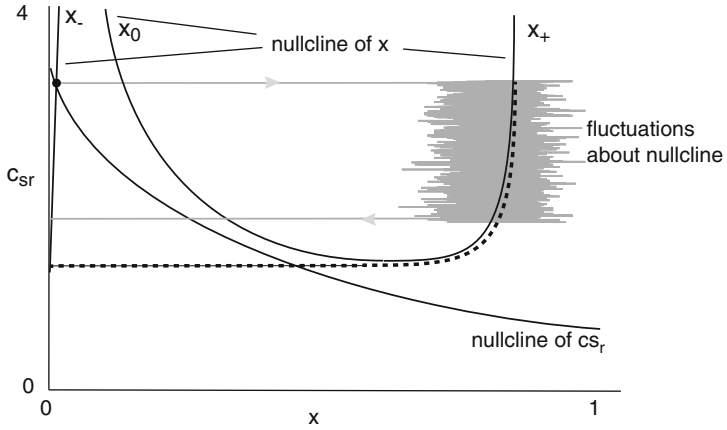


Fig. 3.9: Sketch of nullclines in the deterministic planar Ca^{2+} spark model with x denoting the fraction of open RyRs and c_{SR} the Ca^{2+} concentration in the junctional SR. The c_{sr} nullcline is a monotonically decreasing function $x(c_{sr})$, whereas the x-nullcline is cubic-like with three branches $x_{\pm}(c_{sr})$ and $x_0(c_{sr})$. (Note that the branch $x_-(c_{sr}) \approx 0$; we have moved it away from the vertical axis for the sake of illustration.) In the given diagram there is a single, stable fixed point on the left-hand branch. In the stochastic version of the model a Ca^{2+} spark initiates a jump to the right-hand branch $x_+(c_{sr})$. This is followed by a stochastic trajectory in which the slow variable $c_{sr}(t)$ moves down the nullcline until it undergoes a noise-induced transition back to the left-hand branch before the knee at $x = x_c$. In the deterministic case, the return transition occurs at the knee (*dashed curve*)

approximation when considering the initiation of a Ca^{2+} spark. However, following Ca^{2+} release from the RyRs, the Ca^{2+} concentration c_{sr} slowly changes according to

$$\tau_{sr}\frac{dc_{sr}}{dt} = -c_{sr}x + k_{sr}[c_0 - c_{sr}]. \tag{3.4.16}$$

where $\tau_{sr} \gg \tau_o \gg \tau_D$. The first term on the right-hand side is the loss of Ca^{2+} through the RyRs, whereas the second term is the influx J_{SR} of Ca^{2+} from the NSR with fixed Ca^{2+} concentration c_0 (see Fig. 3.7). The variation of c_{sr} means that one has to modify the analysis of the time to terminate the Ca^{2+} spark. Following Hinch [263], this can be achieved by combining the theory of stochastic transitions with the classical phase-plane analysis of slow–fast excitable systems (see [322] and Sect. 3.5). That is, (3.4.15) and (3.4.16) form an excitable system with the fraction x of open RyRs acting as the fast variable and c_{sr} acting as the slow variable. In Fig. 3.9 we sketch the nullclines of the deterministic system in a parameter regime where there is a single, stable fixed point (x^*, c_{sr}^*). In the full stochastic model, the initiation of a Ca^{2+} spark induces a transition to the right-hand x-nullcline according to $x_-(c_{sr}^*) \to x_+(c_{sr}^*)$. The slow variable then moves down the right-hand nullcline $x_+(c_{sr})$ according to the equation

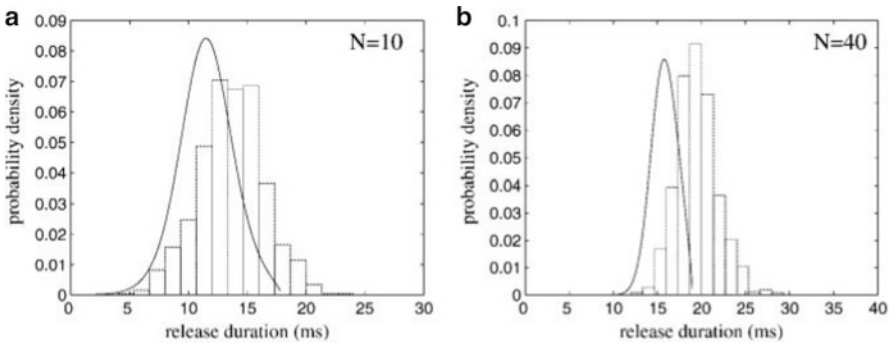


Fig. 3.10: Distribution of spark durations for clusters containing (**a**) 10 RyRs and (**b**) 40 RyRs. The results of Monte Carlo simulations are plotted as a histogram, whereas the *solid line* is the analytical result calculated using stochastic phase-plane analysis. Additional parameters are $c_0 = 3.5$, and $k_{\rm sr} = 0.3$. Adapted from Hinch [263]]

$$\tau_{\rm sr}\frac{dc_{\rm sr}}{dt} = -c_{\rm sr}x_+(c_{\rm sr}) + k_{\rm sr}[c_0 - c_{\rm sr}]. \tag{3.4.17}$$

That is, although x is a stochastic variable, it fluctuates much faster than the dynamics of c_{sr} so one can substitute a time-averaged value of x in (3.4.16).

Suppose that $c_{\rm sr}(t)$ is the solution of (3.4.17) with $c_{\rm sr}(0) = c^*_{sr}$, that is, the Ca^{2+} spark occurs at $t = 0$. In principle, the spark can terminate at any time $t > 0$ due to fluctuations in the number of open RyRs. Using a separation of time scales, we can estimate the rate of transition λ back to the left-hand branch at time t by solving the FPT problem using a diffusion approximation with $c_{sr}(t)$ fixed. Since λ depends on c_{sr}, we have a time-dependent transition rate $\lambda(t) = \lambda(c_{sr}(t))$. One can now calculate the distribution of spark durations T. Let $P(\tau) = {\rm Prob}(T > \tau)$ and introduce the spark duration probability density $p(\tau) = -dP/d\tau$. The probability that a spark terminates in an infinitesimal time interval $\delta\tau$ is $\lambda(\tau)\delta\tau$, so that

$$P(\tau + \delta\tau) = P(\tau)(1 - \lambda(\tau)\delta\tau).$$

Dividing both sides by $\delta\tau$ and taking the limit $\delta\tau \to 0$ gives

$$\frac{dP}{d\tau} = -\lambda(\tau)P(\tau),$$

which can be integrated to yield $P(\tau) = \exp\left(-\int_0^\tau \lambda(t)dt\right)$. Note that by definition $P(0) = 1$. It follows that

$$p(\tau) = \lambda(\tau)\exp\left(-\int_0^\tau \lambda(t)dt\right). \tag{3.4.18}$$

An illustration of the distribution of spark durations is shown in Fig. 3.10; the results are consistent with experimental data of Wang et al. [670].

Finally, note that one of the major simplifications of the Hinch model [263] is that the background Ca^{2+} concentrations in the mytoplasm (c_m) and the network SR (c_{nsr}) are held fixed. It thus fails to capture the collective behavior of a large population of Ca^{2+} release units (CaRUs), which are coupled via global changes in these background concentrations (assuming diffusion can be neglected on the relevant time-scales). This has motivated the development of a whole-cell model of calcium-induced calcium release in cardiac myocytes, based on a system of $\mathcal{N}$ globally-coupled CaRUs [682, 683].

3.5 Membrane Voltage Fluctuations and Spontaneous Action Potentials

Conductance-based models of the Hodgkin–Huxley type have been used to describe many important features of the electrophysiology of neurons and other secretory cells [322]. It is typically assumed that the number of voltage-gated ion channels is sufficiently large so that one can determine the average transmembrane ionic currents based on the opening probabilities of individual channels, which is an application of the law of large numbers. However, the resulting deterministic equations cannot account for spontaneous events driven by ion channel fluctuations. In this section, we describe how to couple the voltage dynamics to the stochastic opening and closing of a finite number of ion channels and show how such fluctuations can initiate a SAP. However, it is first useful to review the classical theory of conductance-based models of neural excitability.

3.5.1 Conductance-Based Model of Neural Excitability

A neuron typically consists of a cell body (or soma) where the nucleus containing DNA is located, a branching output structure known as the *axon* and a branching input structure known as the *dendritic tree* (see Fig. 3.11). Neurons mainly communicate with each other by sending electrical impulses or spikes (action potentials) along their axons. (Some neurons are also coupled diffusively via gap junctions.) These axons make contacts on the dendrites of other neurons via microscopic junctions known as *synapses*. The basic components of synaptic processing induced by the arrival of an action potential are shown in the inset of Fig. 3.11. Depolarization of the presynaptic axon terminal causes voltage-gated Ca^{2+} channels within an active zone to open. The influx of Ca^{2+} produces a high concentration of Ca^{2+} near the active zone, which in turn causes vesicles containing neurotransmitter to fuse with the presynaptic cell membrane and release their contents into the synaptic cleft (exocytosis). The released neurotransmitter molecules then diffuse across the synaptic cleft and bind to specific receptors on the postsynaptic membrane. These receptors cause ion channels to open, thereby changing the membrane conductance and membrane potential of the postsynaptic cell.

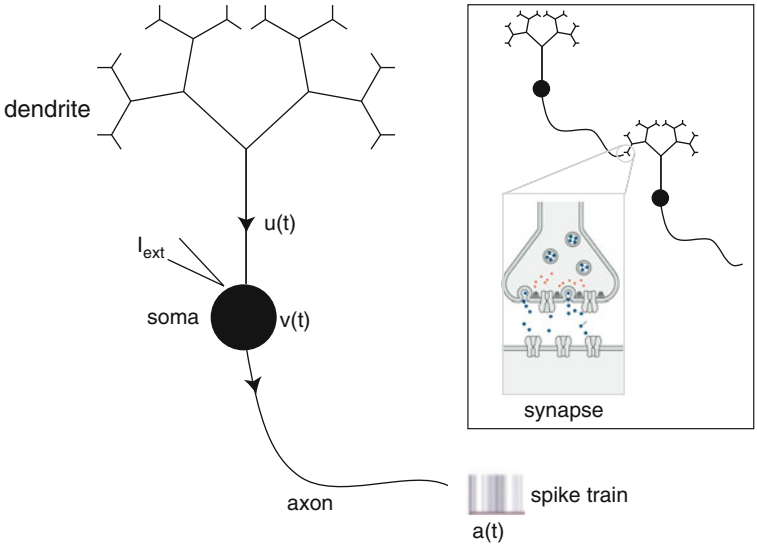


Fig. 3.11: Basic structure of a neuron. [Inset shows a synaptic connection from an upstream or presynaptic neuron and a downstream or postsynaptic neuron.] See text for details

The opening of synaptic ion channels results in the flow of electrical current along the dendritic tree of the stimulated neuron. If the total synaptic current from all of the activated synapses forces the electrical potential within the cell body to cross some threshold, then the neuron fires a spike. The standard biophysical model for describing the dynamics of a single neuron with somatic membrane potential v is based upon conservation of electric charge:

$$C\frac{dv}{dt} = -I_{\text{con}} + u + I_{\text{ext}}, \tag{3.5.1}$$

where C is the cell capacitance, I_{con} is the membrane current, u denotes the sum of synaptic currents entering the cell body, and I_{ext} describes any externally injected currents. Ions can diffuse in and out of the cell through ion-specific channels embedded in the cell membrane. Ion pumps within the cell membrane maintain concentration gradients, such that there is a higher concentration of Na^+ and Ca^{2+} outside the cell and a higher concentration of K^+ inside the cell. The membrane current through a specific channel varies approximately linearly with changes in the potential v relative to some equilibrium or reversal potential, which is the potential at which there is a balance between the opposing effects of diffusion and electrical forces. Summing over all channel types, the total membrane current (flow of positive ions) leaving the cell through the cell membrane is

$$I_{\text{con}} = \sum_s g_s(v - V_s), \tag{3.5.2}$$

where g_s is the conductance due to channels of type s and V_s is the corresponding reversal potential. In the case of a channel selective to a single ion, V_s satisfies the Nernst equation (see Ex. [2.4])

$$V_s = \frac{k_B T}{q} \ln\left(\frac{[outside]_s}{[inside]_s}\right), \tag{3.5.3}$$

where q is the charge of the ion, k_B is the Boltzmann constant, T is temperature (in degrees Kelvin), and $[outside]_s, [inside]_s$ denote the extracellular and intracellular concentrations of the given ion. Typical values for the common ion species are $V_{\mathrm{K}} \approx -75$ mV, $V_{\mathrm{Na}} \approx 50$ mV, $V_{\mathrm{Ca}} \approx 150$ mV, and $V_{\mathrm{Cl}} \approx -60$ mV (which is close to the resting potential of the cell).

The generation and propagation of an action potential arises from nonlinearities associated with active membrane conductances. Recordings of the current flowing through single channels indicate that channels fluctuate rapidly between open and closed states in a stochastic fashion, as described in Sect. 3.1. Nevertheless, most models of a neuron use deterministic descriptions of conductance changes, under the assumption that there are a large number of approximately independent channels of each type. It then follows from the law of large numbers that the fraction of channels open at any given time is approximately equal to the probability that any one channel is in an open state. The conductance g_s for ion channels of type s is thus taken to be the product $g_s = \bar{g}_s P_s$ where $\bar{g}_s$ is equal to the density of channels in the membrane multiplied by the conductance of a single channel and P_s is the fraction of open channels. The voltage dependence of the probabilities P_s in the case of a delayed-rectifier K^+ current and a fast Na^+ current was originally obtained by Hodgkin and Huxley [265] as part of their Nobel prize winning work on the generation of action potentials in the squid giant axon. The delayed-rectifier K^+ current is responsible for terminating an action potential by repolarizing a neuron. One finds that opening of the K^+ channel requires structural changes in 4 identical and independent subunits so that $P_K = n^4$ where n is the probability that any one gate subunit has opened. In the case of the fast Na^+ current, which is responsible for the rapid depolarization of a cell leading to action potential generation, the probability of an open channel takes the form $P_{Na} = m^3 h$ where m^3 is the probability that an activating gate is open and h is the probability that an inactivating gate is open. Depolarization causes m to increase and h to decrease, whereas hyperpolarization has the opposite effect.

The dynamics of the gating variables m, n, h are usually formulated in terms of a simple kinetic scheme that describes voltage-dependent transitions of each gating subunit between open and closed states. More specifically, for each $X \in \{m, n, h\}$

$$\frac{dX}{dt} = \alpha_X(v)(1 - X) - \beta_X(v)X, \tag{3.5.4}$$

where $\alpha_X(v)$ is the rate of the transition $closed \to open$ and $\beta_X(v)$ is the rate of the reverse transition $open \to closed$ [see equation (3.2.5)]. Equation (3.5.4) can be rewritten in the alternative form

$$\tau_X(v)\frac{\mathrm{d}X}{\mathrm{d}t} = X_\infty(v) - X, \qquad \text{with } X \in \{m, n, h\}, \tag{3.5.5}$$

where

$$\tau_X(v) = \frac{1}{\alpha_X(v) + \beta_X(v)}, \qquad X_\infty(v) = \alpha_X(v)\tau_X(v).$$

It follows that the conductance variables m, n, and h approach the asymptotic values $m_\infty(v)$, $n_\infty(v)$, and $h_\infty(v)$ exponentially with time constants $\tau_m(v)$, $\tau_n(v)$, and $\tau_h(v)$, respectively. From basic thermodynamic arguments, the opening and closing rates are expected to be exponential functions of the voltage. Hodgkin and Huxley [265] fitted exponential-like functions to the experimental data obtained from the squid axon:

$$\alpha_m = \frac{0.1(v+40)}{1-\exp[-0.1(v+40)]} \qquad \alpha_h = 0.07\exp[-0.05(v+65)],$$
$$\alpha_n = \frac{0.01(v+55)}{1-\exp[-0.1(v+55)]} \qquad \beta_m = 4.0\exp[-0.556(v+65)],$$
$$\beta_h = \frac{1}{1+\exp[-0.1(v+35)]} \qquad \beta_n = 0.125\exp[-0.125(v+65)].$$

All potentials are measured in mV, all times in ms, and all currents in μA per cm^2. The corresponding asymptotic functions $X_\infty(v)$ and time constants $\tau_X(v)$ are plotted in Fig. 3.12.

We can now write down the Hodgkin–Huxley model for the generation of an action potential, which takes the membrane current to be the sum of a leakage current, a delayed-rectifier K^+ current, and a fast Na^+ current,

$$C\frac{dv}{dt} = f(v,m,n,h) + I_{\text{ext}}, \tag{3.5.6}$$

with

$$f(v,m,n,h) = -\bar{g}_{\text{Na}}m^3h(v-V_{\text{Na}}) - \bar{g}_{\text{K}}n^4(v-V_{\text{K}}) - \bar{g}_{\text{L}}(v-V_{\text{L}}). \tag{3.5.7}$$

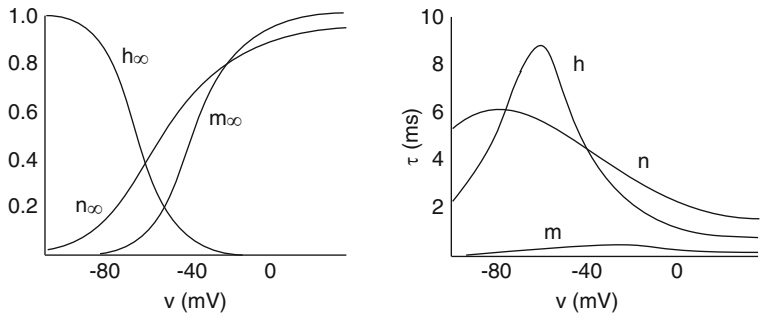


Fig. 3.12: Voltage-dependent steady-state levels of activation and inactivation (*left panel*) and voltage-dependent time constants (*right panel*) for the Hodgkin–Huxley model

The maximal conductances and reversal potentials used in the original model are $\bar{g}_{\mathrm{L}} = 0.003$ mS/mm^2, $\bar{g}_{\mathrm{K}} = 0.36$ mS/mm^2, $\bar{g}_{\mathrm{Na}} = 1.2$ mS/mm^2, $V_{\mathrm{L}} = -54.387$ mV, $V_{\mathrm{K}} = -77$ mV, and $V_{\mathrm{Na}} = 50$ mV. Note that the leakage current groups together various voltage-independent processes such as the currents carried by ion pumps that maintain the concentration gradients. The variables m, n, h evolve according to equation (3.5.4). The temporal evolution of the variables v, f, m, n, h during a single action potential is shown in Fig. 3.13. Injection of a depolarizing current induces a rapid increase in the m variable describing activation of the Na^+ current. Since the slower h variable is initially around 0.6, there is a large influx of Na^+ ions, producing a sharp downward spike in the membrane current and a rapid depolarization through positive feedback. However, the rise in the membrane potential causes the Na^+ conductance to inactivate by driving h towards zero. In addition, the depolarization activates the K^+ conductance, resulting in a subsequent hyperpolarization.

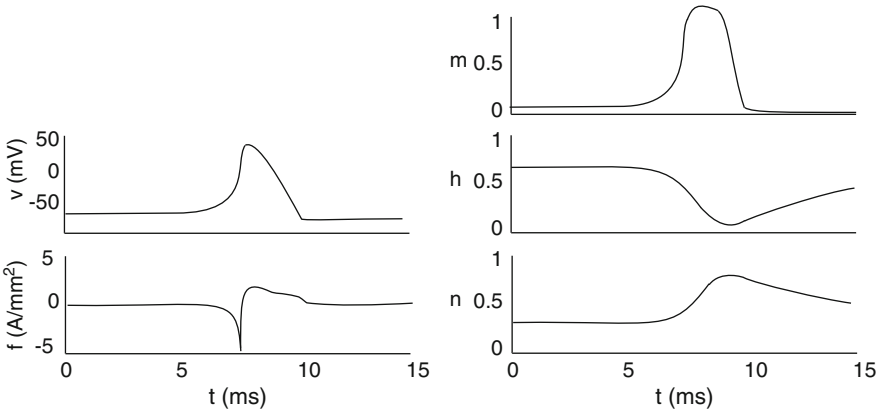


Fig. 3.13: The dynamics of v, f, n, m, h in the Hodgkin–Huxley model during the firing of an action potential induced by a current injection at $t = 5$ ms

3.5.2 Neural Excitability and Phase-Plane Analysis

In order to understand the basic mechanism of neural excitability, we consider a simplified version of the Hodgkin–Huxley model, namely, the Morris–Lecar model [454], which takes the form of a planar dynamical system

$$\frac{dv}{dt} = a(v) f_{\mathrm{Na}}(v) + w f_{\mathrm{K}}(v) - g(v) \equiv f(v, w) \tag{3.5.8a}$$

$$\frac{dw}{dt} = \frac{w_\infty(v) - w}{\tau_w(v)} \equiv \varepsilon g(v, w), \tag{3.5.8b}$$

where w represents the fraction of open K^+ channels, $f_i(v) = g_i(v_i - v)$, $g(v) = g_0(v - v_0)$, and

$$a(v) = \frac{\alpha_{\mathrm{Na}}(v)}{\alpha_{\mathrm{Na}}(v) + \beta_{\mathrm{Na}}(v)}, \quad w_\infty(v) = \frac{\alpha_{\mathrm{K}}(v)}{\alpha_{\mathrm{K}}(v) + \beta_{\mathrm{K}}(v)}.$$

Here

$$\alpha_{\mathrm{Na}}(v) = \mathrm{e}^{4(\gamma_{\mathrm{Na}} v + \kappa_{\mathrm{Na}})}, \beta_{\mathrm{Na}}(v) = 1, \alpha_{\mathrm{K}}(v) = \mathrm{e}^{\gamma_{\mathrm{K}} v + \kappa_{\mathrm{K}}} = \beta_{\mathrm{K}}(v)^{-1}.$$

The fraction of Na^+ channels (or Ca^{2+} channels in the original formulation of the model) is assumed to be in QSS. The generation of action potentials can be analyzed using a slow/fast analysis of the deterministic system, under the assumption that the dynamics of w is slow relative to that of v, that is, $\varepsilon \ll 1$. The fast variable

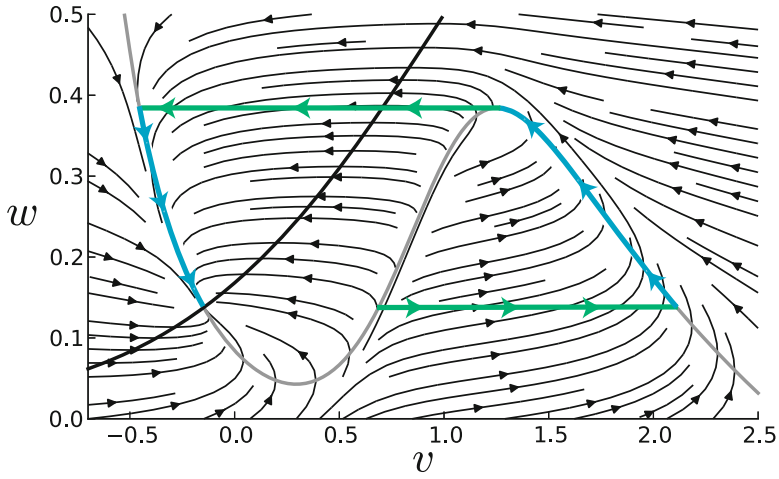


Fig. 3.14: Deterministic phase-plane dynamics of the ML model. Nullclines: $\dot{v} = 0$ (*gray*) and $\dot{w} = 0$ (*black*). *Black streamlines* represent deterministic trajectories. *Green/blue* curves represent an action potential in the limit of slow w

v has a cubic-like nullcline (along which $\dot{v} = 0$) and the slow variable has a monotonically increasing nullcline (along which $\dot{w} = 0$) (see Fig. 3.14). It is assumed that the nullclines have a single intersection point at (v^*, w^*). This corresponds to a fixed point of the system, which we identify with the resting state. A schematic diagram of the phase plane is shown in Fig. 3.15. For a finite range of values of w, there exist three solutions $v = v(w)$ of the equation $f(v,w) = 0$, which we denote by $V_-(w), V_0(w)$, and $V_+(w)$. Whenever these solutions coexist, we have the ordering $V_-(w) \le V_0(w) \le V_+(w)$. Let W_* denote the minimal value of w for which $V_-(w)$ exists and let W^* denote the maximal value of w for which $V_+(w)$ exists.

Suppose that the fixed point is located on the left-hand branch close to the minimum of the cubic. It is straightforward to show that the fixed point is linearly stable

by evaluating the eigenvalues of the corresponding Jacobian. Moreover, the system is excitable in the sense that sufficiently large perturbations of the resting state result in a time-dependent trajectory taking a prolonged excursion through state space before returning to the resting state (see Fig. 3.15). Such a trajectory rapidly transitions to the right branch V_+, after which it slowly moves upward in a neighborhood of the branch before reaching the maximum. It then rapidly transitions back to the left branch V_- followed by slow returns to the resting state along this branch. The time-dependent plot of the variable v can be interpreted as an action potential. Since the resting state is linearly stable, small perturbations simply result in small excursions that decay exponentially in time. Hence, there is effectively a threshold phenomenon in which subthreshold perturbations result in a simple return to the resting state, whereas super-threshold perturbations generate an action potential.

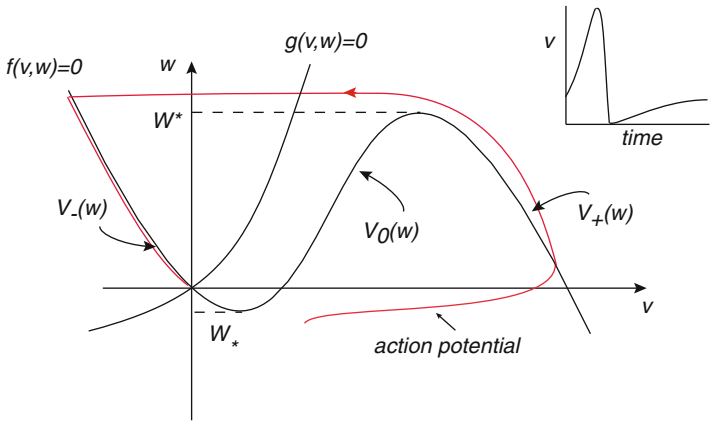


Fig. 3.15: Schematic diagram illustrating the trajectory of a single action potential in the phase plane for the ML equations. The unique rest point is stable. Inset shows the action potential as a function of time

A more mathematical description of the above events can be developed in terms of singular perturbation theory [173, 322, 548]. Due to the separation of time scales with $\varepsilon \ll 1$, the fast variable v rapidly adjusts whenever it can to maintain the quasi-equilibrium $f(v,w)$. This can be captured by introducing the slow time scale $\tau = \varepsilon t$ such that (3.5.8) become

$$\varepsilon \frac{dv}{d\tau} = f(v,w), \quad \frac{dw}{d\tau} = g(v,w). \tag{3.5.9}$$

Now setting $\varepsilon = 0$ and assuming that v is moving along the stable branches $V_\pm(w)$ of $f(v,w) = 0$, the dynamics of the recovery variable reduces to

$$\frac{dw}{d\tau} = g(V_\pm(w),w) \equiv G_\pm(w). \tag{3.5.10}$$

In the case of rapid transitions between the left and right branches, the dynamics with respect to the fast time scale can be approximated by setting $\varepsilon = 0$ in (3.5.8),

$$\frac{dv}{dt} = f(v,w), \quad \frac{dw}{dt} = 0. \tag{3.5.11}$$

Thus, on this time scale, w is constant and v converges to a stable solution of $f(v,w) = 0$. Suppose that the system starts from a super-threshold initial condition (v_0, w_0) such that $v_0 > V_0(w_0)$. After rapidly reaching the right branch, it takes a finite time to reach the upper "knee" of the nullcline $f(v,w)$ and is obtained by integrating (3.5.10):

$$T_e = \int_{w_0}^{W^*} \frac{dw}{G_+(w)}. \tag{3.5.12}$$

On the other hand, the time taken to return to the resting state along the left branch is infinite, since $G_-(w)$ vanishes at the fixed point.

In the above we have focused on so-called type II excitability in which there is a globally attracting resting state. This should be contrasted with another form of excitability exhibited by the ML model, which is called *type I excitability* [173]. From a dynamical systems perspective, the latter is associated with a *saddle node on a limit cycle* (SNLC), also known as a *saddle node on an invariant circle* (SNIC). This is illustrated in Fig. 3.16, which shows three fixed points corresponding to a stable node (the resting state), a saddle, and an unstable node. The stable node and saddle lie on a closed curve consisting of the unstable manifold of the saddle. Under a saddle-node bifurcation the saddle and unstable node annihilate resulting in the

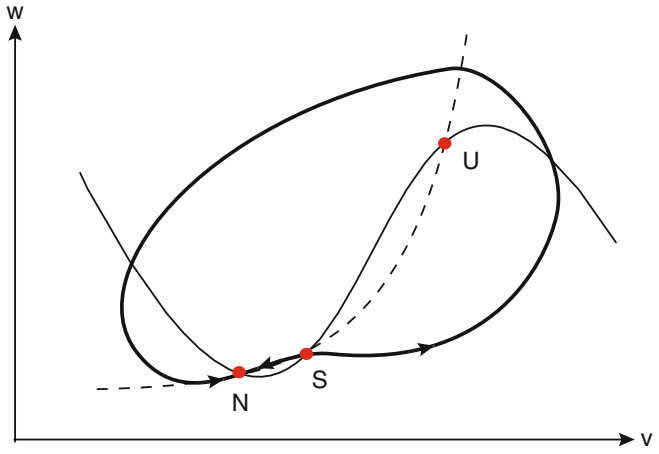


Fig. 3.16: Sketch of type I excitability in the deterministic ML model. Nullclines: $\dot{v} = 0$ (*solid*) and $\dot{w} = 0$ (*dashed*). There are three fixed points: a stable node (N), a saddle (S), and an unstable node (U). The fixed points S and N lie on a closed curve (*thick black*) that is the unstable manifold of S. Under a change of parameters, the fixed points S and N can annihilate via a saddle-node bifurcation resulting in the transition to a limit cycle oscillator

formation of a limit cycle oscillator; this bifurcation scenario is distinct from the onset of oscillations typical of type II excitability, which occur via a Hopf bifurcation (see Box 3B). Below the SNIC bifurcation point, an action potential will occur if the resting state is perturbed beyond the stable manifold of the saddle. In the following we will focus on the effects of stochastic ion channels on type II excitability.

3.5.3 Stochastic Conductance-Based Model

Let us now consider a conductance-based model of a neuron, in which the stochastic opening of ion channels generates a stochastic ionic current that drives the membrane voltage. It is then possible that ion channel noise induces SAPs, which can have a large effect on a neuron's function [678]. If SAPs are too frequent, a neuron cannot reliably perform its computational role. Hence, ion channel noise imposes a fundamental limit on the density of neural tissue. Smaller neurons must function with fewer ion channels, making ion channel fluctuations more significant and more likely to cause a SAP. In order to investigate SAPs, we will consider a stochastic version of the ML model [476] consisting of N sodium and M potassium channels:

$$\frac{dV}{dt} = F(V,m,n) \equiv \frac{n}{N} f_{\rm Na}(V) + \frac{m}{M} f_{\rm K}(V) - g(V). \tag{3.5.13}$$

We assume that each channel can either be open or closed and can switch between each state according to the kinetic scheme

$$C \underset{\beta_i(v)}{\overset{\alpha_i(v)}{\rightleftarrows}} O, \quad i = \text{Na, K}. \tag{3.5.14}$$

(Note that a more detailed biophysical model would need to treat each ion channel as a cluster of subunits rather than a single unit [224]. In other words, the Markov chain of events associated with opening and closing of an ion channel would involve transitions between more than two internal states; see Sect. 3.1.)

The Na^+ channels open and close rapidly relative to the voltage and K^+ dynamics. We are particularly interested in how fluctuations affect the initiation of an action potential due to the opening of a finite number of Na^+ channels. Therefore, we imagine freezing the slow K^+ channels, so that they effectively act as a leak current, and simplify the sodium channels by treating each as a single activating subunit. The stochastic membrane voltage then evolves according to the piecewise deterministic equation

$$\frac{dV}{dt} = F(V,n) \equiv \frac{n}{N} f(V) - g(V), \tag{3.5.15}$$

where $f(V) = f_{\rm Na}(V)$, $g(V) = -g_{\rm eff}[V_{\rm eff} - V] - I$ represents the sum of effective leakage currents and external inputs I, and $n(t)$ is the number of open sodium

channels. The right-hand side of (3.5.15) is negative for large V and positive for small V, which implies that the voltage V is confined to some bounded domain $[V_1, V_2]$. Note that (3.5.15) only holds between jumps in the number of open ion channels, with the latter described by the master equation (3.2.1). The stochastic process defined by (3.5.14) and (3.5.15) is an example of a *stochastic hybrid system* with piecewise deterministic dynamics. There has been a lot of recent interest in such systems, particularly within the context of conductance-based models [74, 81, 321, 476, 491, 600, 663]. They also arise in models of polymerization (Sect. 4.1), biochemical chemotaxis (Sect. 5.3), gene networks (Sect. 6.4), and active motor transport (Sect. 7.4). The associated probability density $p(v,n,t)$, which is defined according to

$$p(v,n,t)dv = \text{Prob}[n(t) = n, v \le V(t) \le v + dv],$$

given an initial condition $V(0) = V_0, n(0) = n_0$, satisfies the *differential Chapman–Kolmogorov (CK) equation*

$$\begin{aligned}\frac{\partial p}{\partial t} = &-\frac{\partial}{\partial v}\left[\left(\frac{n}{N}f(v) - g(v)\right)p\right] + \omega_+(v,n-1)p(v,n-1,t) \\ &+\omega_-(v,n+1)p(v,n+1,t) - [\omega_+(v,n) + \omega_-(v,n)]p(v,n,t),\end{aligned} \tag{3.5.16}$$

with

$$\omega_+(v,n) = \alpha(v)(N-n), \quad \omega_-(v,n) = \beta(v)n. \tag{3.5.17}$$

The first term on the right-hand side of (3.5.16) generates the deterministic dynamics for a given n – see the discussion of the Liouville equation in Sect. 2.2.1 – whereas the other terms represent the transitions between discrete states of the Markov chain. If the continuous process were an SDE, then there would be an additional n-dependent diffusion term on the right-hand side. Techniques for analyzing differential CK equations will be developed in Sect. 7.4, 7.6, and 10.1.

In order to investigate action potential initiation, we will make the following approximations:

(i) For sufficiently large N, we approximate the jump Markov process for the ion channels by a continuous Markov process using a diffusion approximation as outlined in Sect. 3.2;
(ii) The transitions between different discrete states are much faster than the voltage dynamics so that, for fixed v, the number of open ion channels is close to the quasi-equilibrium $x^* = a(v) \equiv \alpha(v)/(\alpha(v) + \beta(v))$. This limiting case was originally considered by Chow and White [116].

Under approximation (i), the voltage dynamics can be described by an SDE of the form (see (3.2.20))

$$dV = \left[f(V)(a(V) + Y(t)/\sqrt{N}) - g(V)\right]dt, \quad dY = -k(V)Y\,dt + b(V)dW(t), \tag{3.5.18}$$

with

$$k(V) = \alpha(V) + \beta(V), \quad b(V) = \sqrt{\frac{2\alpha(V)\beta(V)}{\alpha(V)+\beta(V)}}.$$

Thus the stochastic voltage is coupled to a fast Ornstein–Uhlenbeck process $Y(t)$. In the case of fast sodium channel kinetics (approximation (ii)), we have $\alpha, \beta = O(1/\varepsilon)$ for some dimensionless parameter $0 < \varepsilon \ll 1$. (The time scale is fixed by taking the relaxation time of the voltage to be of $O(1)$.) It follows that $k(V) = O(1/\varepsilon)$ and $b(V) = O(1/\sqrt{\varepsilon})$. Hence, we can take $Y(t)$ to be in quasi-equilibrium for a given V (fast ion channels), that is, $Y(t)dt \approx k(V)^{-1} b(V) dW(t)$. This then yields a scalar SDE for the voltage:

$$dV = [f(V)a(V) - g(V)]\,dt + \frac{1}{\sqrt{N}}\sigma(V)f(V)dW(t), \tag{3.5.19}$$

where

$$\sigma(V) = \frac{b(V)}{k(V)} = \frac{1}{\alpha(V)+\beta(V)}\sqrt{\frac{2\alpha(V)\beta(V)}{\alpha(V)+\beta(V)}}. \tag{3.5.20}$$

In deriving (3.5.19), we have effectively taken a zero correlation limit of an Ornstein–Uhlenbeck process. It can be shown that the multiplicative noise term should be interpreted in the sense of Stratonovich [204, 321]. However, for large N this yields an $O(1/N)$ correction to the drift term in the FP equation, which can be dropped. We thus obtain the FP equation

$$\frac{\partial p(v,t)}{\partial t} = -\frac{\partial}{\partial v}[A(v)p(v,t)] + \frac{1}{2N}\frac{\partial^2}{\partial v^2}[B(v)p(v,t)], \tag{3.5.21}$$

with

$$A(v) = f(v)a(v) - g(v), \quad B(v) = [\sigma(v)f(v)]^2. \tag{3.5.22}$$

The FP equation is supplemented by reflecting boundary conditions at $v = V_1, V_2$:

$$J(V_1,t) = J(V_2,t) = 0, \tag{3.5.23}$$

with

$$J(v,t) = A(v)p(v,t) - \frac{1}{2N}\frac{\partial}{\partial v}B(v)p(v,t). \tag{3.5.24}$$

Note that equation (3.5.21) is identical in form to (3.2.10) except we now have a stochastic process with respect to membrane voltage rather than fraction of open ion channels; the latter is slaved to the voltage.

A key property that one would like to calculate is the mean time to fire an action potential (MFPT) as a function of the stimulus current I. In the absence of noise, the system evolves according to the deterministic equation

$$\frac{dv}{dt} = A(v) = \frac{\alpha(v)}{\alpha(v)+\beta(v)}f(v) - g(v) \equiv -\frac{dU(v)}{dv}, \tag{3.5.25}$$

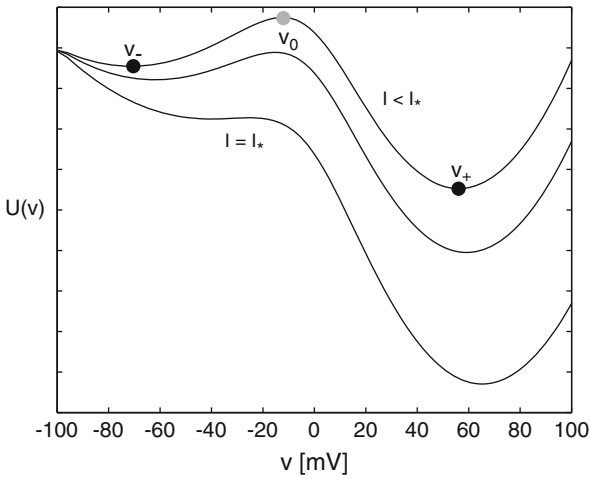


Fig. 3.17: Plot of deterministic potential $U(v)$ as a function of voltage v for different values of the external stimulus current I. Parameter values are $N = 10$, $v_{\rm Na} = 120$ mV, $v_{\rm eff} = -62.3$ mV, $g_{\rm Na} = 4.4$ ms/cm^2, $g_{\rm eff} = 2.2$ ms/cm^2, $\beta = 0.8$ s^{-1}, and $\alpha(v) = \beta \exp[(v+1.2)/9]$

where $U(v)$ is a deterministic potential. In Fig. 3.17, we plot $U(v)$ as a function of v for various values of the external input current and the particular transition rates

$$\alpha(v) = \beta \exp\left(\frac{2(v-v_1)}{v_2}\right), \quad \beta = \text{constant}.$$

The minima and maxima of the potential correspond to stable and unstable fixed points of the deterministic dynamics, respectively. It can be seen that below a threshold current I_*, $I < I_*$, there exist two stable fixed points $v_\pm$ (minima) separated by an unstable fixed point at v_0 (maximum), that is, the system exhibits bistability. The left-hand fixed point represents the resting state, whereas the right-hand fixed point corresponds to an excited state. Thus, in the bistable regime the deterministic system requires an external perturbation in order to generate an action potential starting from the resting state. On the other hand, for the stochastic system it is possible that fluctuations in the opening and closing of Na^+ ion channels induce a transition from the resting state to the excited state by crossing over the potential hill at v_0. Of course, once such an event occurs, one has to take into account the K^+ dynamics in order to incorporate the effects of repolarization that return the system to the resting state. If one includes the slow opening and closing of these channels, then the underlying deterministic system becomes excitable rather than bistable. For simplicity, we will assume that this does not significantly affect the noise-induced initiation of an action potential. It turns out that such an assumption breaks down if fluctuations in the opening and closing of K^+ channels become significant [476] (see Sect. 10.1).

The calculation of the mean time to escape from the resting state using the diffusion approximation proceeds along identical lines to Sect. 3.3. After taking $v \to x$, we obtain equation (3.3.34) with the FP quasi-potential $\Psi \equiv \Phi_{\rm FP}$ determined by equations (3.2.16) and (3.5.22):

$$\Phi_{\rm FP}(x) = -2\int^x \frac{A(y)}{B(y)}dy = -2\int^x \frac{f(y)a(y)-g(y)}{\sigma(y)f(y)}dy. \tag{3.5.26}$$

Keener and Newby [321] explicitly calculated the MFPT and compared it with Monte Carlo simulations of the full stochastic model whose probability density evolves according to the CK equation (3.5.17). A summary of their findings is shown schematically in Fig. 3.18. The main observation is that although the Gaussian-like diffusion approximation does well in the superthreshold regime ($I > I_*$), it deviates significantly from the full model results in the subthreshold regime ($I < I_*$), where it overestimates the mean time to spike. This is related to the fact that the

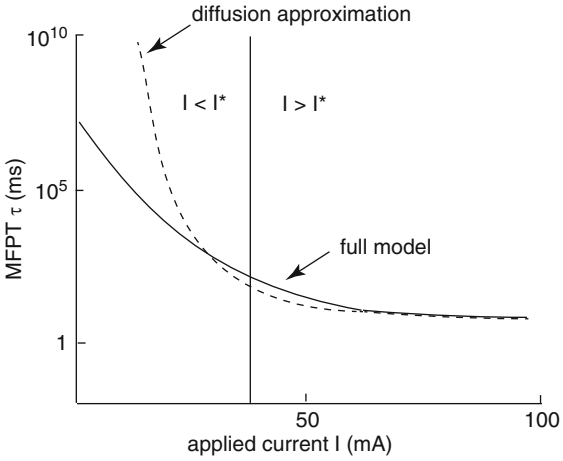


Fig. 3.18: Schematic diagram comparing MFPT calculated using the diffusion approximation with the MFPT of the full system obtained using Monte Carlo simulations. The scales of the axes are based on numerical results carried out in [321] for $N = 10$

quasi-potential of the steady-state density under the diffusion approximation generates exponentially large errors in the MFPT. Thus more advanced asymptotic methods are required such as the WKB approximation [321]. The application of WKB methods to FPT problems is developed in Chap. 10.

3.6 Stochastic Gating Model of Confinement

An interesting problem is the escape of particles from a bounded domain via the random opening and closing of one or more channels embedded in the boundary of the domain (see Fig. 3.19). One important application is to modeling the escape of diffusing proteins from a corral in the plasma membrane (see Sect. 7.1). Consider, in particular, the spatially homogeneous stochastic gating model of Brown and collaborators [79, 380], in which diffusion within the domain is relatively fast so that the molecules are well mixed and one can ignore spatial effects. Let $P_n(t)$ be the probability that there are n free particles within the domain at time t. Denote the state of the stochastic gate at time t to be the binary random variable $\mu(t)$ with $\mu(t) = 1$ ($\mu(t) = 0$) corresponding to the open (closed) state. The opening and closing of the stochastic gate is governed by the two-state Markov process

$$\begin{aligned} \frac{d\mathscr{P}_1}{dt} &= -\gamma_- \mathscr{P}_1 + \gamma_+ \mathscr{P}_0, \\ \frac{d\mathscr{P}_0}{dt} &= \gamma_- \mathscr{P}_1 - \gamma_+ \mathscr{P}_0, \end{aligned} \tag{3.6.1}$$

where $\mathscr{P}_1(t)$ and $\mathscr{P}_0(t)$) are the probabilities that the gate is open and closed, respectively, at time t, and $\gamma_\pm$ are the transition rates between the two states. Particles can only transfer between the exterior and interior of the domain when the gate is open, in which case the rates of outflux and influx are γ and k. Note that k will depend linearly on the concentration of particles outside the domain. The probability distribution $P_n(t)$ evolves according to the nonautonomous birth–death master equation

$$\frac{dP_n}{dt} = \mu(t)\left[kP_{n-1}(t) + (n+1)\gamma P_{n+1}(t) - (k+\gamma n)P_n(t)\right] \tag{3.6.2}$$

with $n \geq 0$ and $P_{-1}(t) \equiv 0$. The positive terms on the right-hand side represent the various transitions into the state (n) whereas the negative terms represent the various transitions from the state (n). The initial condition is $P_n(0) = \delta_{n,n_0}$; i.e., at time $t = 0$ there are n_0 free particles within the domain. First, suppose that the gate is always open so that (3.6.2) reduces to the autonomous master equation

$$\frac{dP_n}{dt} = kP_{n-1}(t) + (n+1)\gamma P_{n+1}(t) - (k+\gamma n)P_n(t). \tag{3.6.3}$$

The mean concentration of particles in the domain, $x(t) = \langle n(t)\rangle/V$, where V is the volume of the domain, evolves according to the simple kinetic equation

$$\frac{dx}{dt} = k/V - \gamma x.$$

This has a steady-state solution $n = Vx = k/\gamma$.

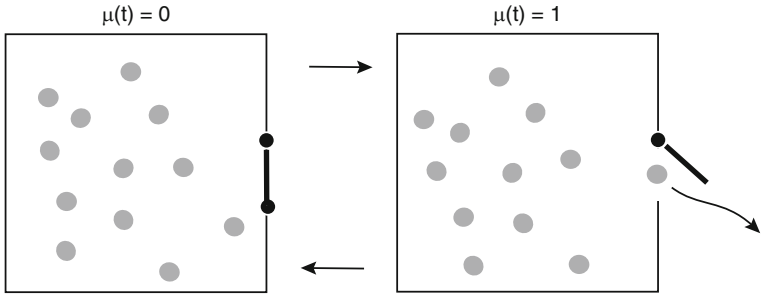


Fig. 3.19: Escape from a domain with a single stochastic gate

Equation (3.6.3) is a rare example of a master equation that can be solved exactly, and one finds that $P_n(t)$ is given by a Poisson distribution. The simplest way to see this is to introduce the generating function

$$G(z,t) = \sum_{m\geq 0} z^m P_m(t),$$

and substitute into (3.6.3):

$$\frac{\partial G}{\partial t} + \gamma(z-1)\frac{\partial G}{\partial z} = k(z-1)G.$$

This is a linear first-order PDE with nonconstant coefficients. A standard method for solving such equations is the *method of characteristics* [554]. The basic idea is to construct characteristic curves $z = z(t)$ along which $G(t) \equiv G(z(t),t)$ satisfies

$$\frac{dG}{dt} = \frac{\partial G}{\partial t} + \frac{dz}{dt}\frac{\partial G}{\partial z},$$

such that the evolution of G is consistent with the original PDE. This then yields the characteristic equations

$$\frac{dz}{dt} = \gamma(z-1), \quad \frac{dG}{dt} = k(z-1)G.$$

Solving for $z(t)$,

$$z(t) = 1 + se^{\gamma t}$$

where s parameterizes the initial data. Then

$$\frac{dG}{dt} = kse^{\gamma t}G, \quad G(t) = F(s)\exp\left(kse^{\gamma t}/\gamma\right)$$

for some function F determined by the initial data. In order to determine the solution $G(z,t)$ we eliminate s in terms of z, which gives

$$G(z,t) = F([z-1]e^{-\gamma t})\exp\left(k(z-1)/\gamma\right). \tag{3.6.4}$$

Since $G(1,t)=1$, we require $F(0)=1$. Moreover, given the initial condition $P_n(0)=\delta_{n,n_0}$, we have $G(z,0)=z^{n_0}$ and $F(z)=(1+z)^{n_0}\mathrm{e}^{-kz/\gamma}$. It follows that

$$G(z,t)=[1+\mathrm{e}^{-\gamma t}(z-1)]^{n_0}\mathrm{e}^{k\left(1-\mathrm{e}^{-\gamma t}\right)(z-1)/\gamma}. \tag{3.6.5}$$

Note that the method of characteristics provides an alternative method for solving equation (2.2.20) for the characteristic function of the Ornstein–Uhlenbeck process (see Ex. 3.10). More importantly, the method can also be applied to nonlinear first-order equations of the general form

$$a(x,t,p)\frac{\partial p}{\partial t}+b(x,t,p)\frac{\partial p}{\partial x}=c(x,t,p)p, \tag{3.6.6}$$

for sufficiently smooth functions a,b,c. These so-called quasilinear equations crop up in many models of transport processes and support nontrivial solutions such as shock waves and rarefaction waves [554]. An application to molecular motors will be developed in Sect. 7.5.

Returning to the stochastic gating model, we Taylor expand $G(z,t)$ in powers of z and find that for $n_0=0$ (bounded domain is initially empty), we have (see Ex. 3.11),

$$P_n(t)=\mathrm{e}^{-\lambda(t)}\frac{\lambda(t)^n}{n!},\quad \lambda(t)=\frac{k}{\gamma}(1-\mathrm{e}^{-\gamma t}), \tag{3.6.7}$$

which is a time-dependent Poisson distribution of rate $\lambda(t)$. It immediately follows that

$$\langle n(t)\rangle=\lambda(t),\quad \mathrm{var}[n(t)]=\lambda(t).$$

In the more general case $n_0\neq 0$, the mean and variance can be calculated from the formulae

$$\langle n(t)\rangle=\left.\frac{\partial G(z,t)}{\partial z}\right|_{z=1},\quad \langle(n^2(t)-n(t))\rangle=\left.\frac{\partial^2 G(z,t)}{\partial z^2}\right|_{z=1}.$$

Calculating these derivatives yields

$$\langle n(t)\rangle=(n_0-k/\gamma)\mathrm{e}^{-\gamma t}+k/\gamma,\ \mathrm{var}[n(t)]=\langle n(t)\rangle-n_0\mathrm{e}^{-2\gamma t}.$$

Let us now turn to the full stochastic gating model, in which the state of the gate is given by the stochastic variable $\mu(t)$ so that there are two levels of stochasticity: the stochastic process of exchange of particles when the gate is open and the random opening and closing of the gate itself. For a given realization $\omega\in\Omega$ of the stochastic gate (a given trajectory $\mu(t)$ through state space), we can repeat the analysis of the autonomous master equation (3.6.3), except that

$$\mathrm{e}^{-\gamma t}\to\mathscr{N}(t)\equiv\mathrm{e}^{-\gamma\int_0^t\mu(t')dt'}.$$

It follows that different realizations of $\mu(t)$ will yield different values of the mean and variance. Hence, a more useful characterization of the statistics is obtained by averaging $\mathscr{N}(t)$ with respect to all possible stochastic realizations of the gate, which is denoted by $\langle \mathscr{N} \rangle_\mu$. The latter can be performed using a method originally developed by Kubo [358] in the study of spectral line broadening in a quantum system and subsequently extended to chemical rate processes with dynamical disorder by Zwanzig [707]. Following Kubo [358], consider the SDE

$$\frac{dX}{dt} = -\gamma \mu(t) X(t),$$

where $\mu(t)$ is a discrete random variable that switches between $\mu = 1$ and $\mu = 0$ according to (3.6.1). This is a simple example of a stochastic hybrid system (see Sect. 3.5). Introduce the probability densities $p_l(x,t)$ with $p_l(x,t)dx = \text{Prob}[\mu(t) = l, x \le X(t) \le x + dx]$, $l = 0,1$, and initial conditions

$$p_l(x,0) = \delta(x-1)\Pi_l.$$

Here Π_l, $l = 0,1$, are the stationary probability distributions of the two-state Markov process (3.6.1):

$$\Pi_1 = \frac{\gamma_+}{\gamma_+ + \gamma_-}, \quad \Pi_0 = \frac{\gamma_-}{\gamma_+ + \gamma_-}.$$

These densities evolve according to the equation

$$\frac{\partial p_1}{\partial t} = \gamma \frac{\partial (x p_1)}{\partial x} - \gamma_- p_1 + \gamma_+ p_0 \tag{3.6.8a}$$

$$\frac{\partial p_0}{\partial t} = \gamma_- p_1 - \gamma_+ p_0. \tag{3.6.8b}$$

We now make the observation that $p(x,t) = p_0(x,t) + p_1(x,t)$ is the probability density for the stochastic process $X(t)$, which has the formal solution

$$X(t) = \mathscr{N}(t) \equiv \mathrm{e}^{-\gamma \int_0^t \mu(t')dt'}$$

together with the constraint that the initial state of the gate $\mu(0)$ is a random variable distributed according to the stationary distributions $\Pi_{0,1}$. Thus, finding the mean of $\mathscr{N}(t)$ with respect to the stochastic process $\mu(t)$ is equivalent to finding the conditional means

$$m_l(t) = \int_0^\infty x p_l(x,t)dx, \quad l = 0,1,$$

and setting

$$\langle \mathscr{N}(t) \rangle_\mu = m_0(t) + m_1(t).$$

In order to determine $m_{0,1}(t)$, take first moments of equations (3.6.8a,b). This yields the matrix equation

$$\frac{d}{dt}\begin{pmatrix} m_1(t) \\ m_0(t) \end{pmatrix} = -\mathscr{A}\begin{pmatrix} m_1(t) \\ m_0(t) \end{pmatrix}, \quad \mathscr{A} = \begin{pmatrix} \gamma_- + \gamma & -\gamma_+ \\ -\gamma_- & \gamma_+ \end{pmatrix}, \tag{3.6.9}$$

which has the solution

$$\begin{pmatrix} m_1(t) \\ m_0(t) \end{pmatrix} = \mathrm{e}^{-t\mathscr{A}}\begin{pmatrix} \Pi_1 \\ \Pi_0 \end{pmatrix}.$$

A similar analysis can be carried out for second moments. One thus finds that the μ-averaged mean and variance are

$$\langle n\rangle_\mu = (n_0 - k/\gamma)\langle \mathscr{N}\rangle_\mu + k/\gamma, \tag{3.6.10}$$

$$\mathrm{var}_\mu[n] = \langle n\rangle_\mu - n_0\langle \mathscr{N}^2\rangle_\mu + (n_0 - k/\gamma)^2\left(\langle \mathscr{N}\rangle_\mu - \langle \mathscr{N}\rangle_\mu^2\right), \tag{3.6.11}$$

where

$$\langle \mathscr{N}(t)^q\rangle_\mu = \begin{pmatrix} 1 \\ 1 \end{pmatrix}^T \exp\left[-t\begin{pmatrix} \gamma_- + q\gamma & -\gamma_+ \\ -\gamma_- & \gamma_+ \end{pmatrix}\right]\begin{pmatrix} \Pi_1 \\ \Pi_0 \end{pmatrix} \tag{3.6.12}$$

for $q = 1, 2$. The averages $\langle \mathscr{N}^q\rangle$, $q = 1, 2$, approach zero as time increases; hence the steady-state mean and variance are both equal to k/γ. There have been a number of extensions of the stochastic gating model. These include taking into account the effects of membrane proteins binding to scaffolding proteins within a corral [68] and the analysis of the narrow escape problem for a particle that can switch between different conformational states and can only exit a domain in one of these states [531].

3.7 Exercises

Problem 3.1 (Chain of ion channel states). The time course of the opening and closing of some ion channels seems to follow a power law rather than an exponential law at large times. One way to understand such power law behavior is to consider an ion channel with N closed states such that the transition to an open state can only take place from state 1 at one end of a chain

$$0 \overset{\alpha}{\leftarrow} 1 \underset{\beta_1}{\overset{\gamma_1}{\rightleftharpoons}} 2 \ldots \underset{\beta_{N-1}}{\overset{\gamma_{N-1}}{\rightleftharpoons}} N.$$

(a) Write down the corresponding set of kinetic equations. Hence, show that when $\gamma_n = \beta_n = 1$ for all n and $\alpha = 1$, we obtain the discrete diffusion equation along a chain with a reflecting boundary at $n = N$ and an absorbing boundary at $n = 0$:

$$\frac{dp_1}{dt} = p_2 - 2p_1,$$
$$\frac{dp_n}{dt} = p_{n-1} + p_{n+1} - 2p_n, \quad 1 < n < N,$$
$$\frac{dp_N}{dt} = p_{N-1} - p_N,$$

where $p_n(t)$ is the probability that the channel is in state n at time t.

(b) Given the initial condition $p_n(0) = \delta_{n,1}$, show that in the large N limit, the exact solution is

$$p_n(t) = \mathrm{e}^{-2t}[I_{n-1}(2t) - I_{n+1}(2t)],$$

where $I_n(z)$ is the modified Bessel function of integer order:

$$I_n(z) = \int_{-\pi}^{\pi} \mathrm{e}^{ink} \mathrm{e}^{z\cos(k)} \frac{dk}{2\pi}.$$

Hint: use discrete Fourier transforms to solve the discrete diffusion equation on the infinite lattice (Box 2A) and then use the method of images to write down the solution for a semi-infinite lattice (by analogy with Ex. 2.12).

(c) When $2t \gg n$, the modified Bessel function has the asymptotic expansion

$$I_n(2t) = \frac{\mathrm{e}^{2t}}{\sqrt{4\pi t}} \left[1 - \frac{4n^2 - 1}{16t} + \ldots\right].$$

Use this to show that, for large t,

$$p_n(t) \approx \frac{n}{2\pi^{1/2} t^{3/2}}.$$

(d) Define $F(t)$ to be the total probability of finding the system in a closed state:

$$F(t) = \sum_{n=1}^{\infty} p_n(t).$$

Show that $dF/dt = -p_1$ and, hence, $F(t) \approx (\pi t)^{-1/2}$ for large N, t.

Problem 3.2 (Ion channel with multiple subunits). Consider an ion channel with k identical open subunits, each of which can be open or closed, and a current only passes if all k subunits are open.

(a) Let S_j denote the state in which j subunits are open and let α, β denote the rates of opening and closing of a single subunit. Write down the corresponding reaction scheme

(b) Derive the kinetic equations for x_j, which is the fraction of channels in state j such that $\sum_{j=0}^{k} x_j = 1$.

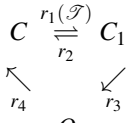
(c) By direct substitution, show that

$$x_j = \frac{k!}{(k-j)!j!} n^j (1-n)^{k-j}$$

is an invariant manifold of the dynamics, provided that

$$\frac{dn}{dt} = \alpha(1-n) - \beta n.$$

Problem 3.3 (Ligand-gated ion channel). Consider the following second-order kinetic scheme for a ligand-gated ion channel:



Here C and C_1 are the closed forms of the receptor, O is the open (conducting) form, and the r_i are voltage-independent transition rates. The transition rate r_1 for $C \to C_1$ depends on the concentration of ligand $\mathscr{T}$. Suppose that we make the following approximations: (i) the transmitter concentration $\mathscr{T}$ occurs as a pulse $\delta(t-t_0)$ for a release event occurring at time $t = t_0$, that is, $r_1(\mathscr{T}) = r_1\delta(t-t_0)$; (ii) the fraction of channels in state C is taken to be fixed at unity—this is reasonable if the number of channels in state C is much larger than the number of channels in states C_1 or O.

(a) Write down the kinetic equations for the fraction of receptors in the states C_1 and O, which are denoted by z and s, respectively.
(b) Solve the resulting pair of inhomogeneous linear equations assuming that $z(0) = s(0) = 0$. In particular show that the fraction of open channels is given by

$$s(t) = r_1 r_3 \left(\frac{1}{\tau_2} - \frac{1}{\tau_1}\right)^{-1} (\mathrm{e}^{-(t-t_0)/\tau_1} - \mathrm{e}^{-(t-t_0)/\tau_2}), \quad t > t_0,$$

with $\tau_1 = 1/(r_2 + r_3)$, $\tau_2 = 1/r_4$.
(c) Show that in the limit $\tau_2 \to \tau_1 = \tau_s$ this reduces to the so-called alpha function

$$s(t) = r_1 r_3 (t-t_0)\mathrm{e}^{-(t-t_0)/\tau_s}, \quad t > t_0.$$

Such a response function is often used to model the response of synaptic receptors following the release of neurotransmitter.

Problem 3.4 (MWC model). Generalize the MWC model of a ligand-gated ion channel from two binding sites to n binding sites. Let R_j, T_j, $j = 0, 1, \ldots, n$, denote the global states (open or closed) with j ligands bound to the allosteric site. Also define K_T and K_R to be the equilibrium constants for binding of an acetylcholine

molecule to an individual site of a closed and open receptor, respectively. Finally, take Y_j to be the equilibrium constant associated with the opening and closing of a receptor with j occupied sites. Defining the fraction of open receptors according to

$$p_{\text{open}} = \frac{\sum_{j=0}^{n}[R_j]}{\sum_{j=0}^{n}[R_j] + \sum_{j=0}^{n}[T_j]}$$

use the law of mass action to derive the sigmoidal function

$$p_{\text{open}} = \frac{Y_0(1+K_R[L])^n}{Y_0(1+K_R[L])^n + (1+K_T[L])^n}.$$

Hint: care needs to be taken in working out the combinatorial factors multiplying K_T and K_R in the reaction diagram.

Problem 3.5 (MWC model and the Boltzmann–Gibbs distribution). Rederive the result for the MWC model in Ex. 3.4 using the Boltzmann–Gibbs distribution (Sect. 1.4). Use the following observations. A microstate is specified by the number of occupied binding sites m, $0 \le m \le n$, and whether the channel is open or closed. The free energy of a given microstate is

$$E_m^r = m\left[\varepsilon_b^r - \mu_0 - k_B T \ln(c/c_0)\right],$$

where ε_b^r is the binding energy when the channel is open $(r = R)$ or closed $(r = T)$ and the chemical potential $\mu = \mu_0 - k_B T \ln(c/c_0)$ takes into account that ligands are being taken out of solution (see Sect. 1.4). Evaluate the partition function

$$Z = \sum_{r=R,T} \frac{n!}{m!(n-m)!} \mathrm{e}^{-E_m^r/k_B T},$$

and hence determine p_{open}. Explain the presence of the combinatorial factor in the definition of Z.

Problem 3.6 (Master equation for an ensemble of ion channels). Consider the master equation for the two-state ion channel model:

$$\frac{d}{dt}P(n,t) = \alpha(N-n+1)P(n-1,t) + \beta(n+1)P(n+1,t) - [\alpha(N-n)+\beta n]P(n,t).$$

(a) By multiplying both sides by n and summing over n, derive the following kinetic equation for the mean $\bar{n} = \sum_{n=0}^{N} nP(n,t)$:

$$\frac{d\bar{n}}{dt} = \alpha(N-\bar{n}) - \beta\bar{n}.$$

(b) Derive a corresponding equation for the variance $\sigma^2 = \langle n^2\rangle - \langle n\rangle^2$. That is, multiply both sides of the master equation by n^2 and sum over n to determine an

equation for the second moment and then use part (a). Show that the variance decays exponentially at a rate $2(\alpha+\beta)$ to the steady-state value

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2}$$

and hence deduce that fluctuations become negligible in the large N limit.

(c) Compare the results obtained from the master equation with the analysis based on the linear noise approximation.

(d) Construct the master equation for an ensemble of N identical, independent channels each of which has two subunits. That is, determine an equation for the evolution of the probability distribution $P_{n_0,n_2}(t)$ that there are n_j ion channels with j open subunits such that $N = n_0 + n_1 + n_2$.

Problem 3.7 (Bistability in an autocatalytic reaction). Consider the following nonlinear, autocatalytic reaction scheme for a protein that can exists in two states X and Y:

$$X \underset{k_2}{\overset{k_1}{\rightleftharpoons}} Y, \quad X + 2Y \xrightarrow{k_3} 3Y.$$

Let $[X]$ and $[Y]$ denote the concentrations of the molecule in each of the two states such that $[X]+[Y] = Y_{\text{tot}}$ fixed. The kinetic equation for $[Y]$ is

$$\frac{d[Y]}{dt} = -k_2[Y] + k_1[X] + k_3 V^2 [Y]^2 [X],$$

where V is cell volume.

(a) Let $y = [Y]/Y_{\text{tot}}$. Show that after an appropriate rescaling of time, the corresponding kinetic equation for y is

$$\frac{dy}{dt} = y(\mu(1-y)y - 1) + \lambda(1-y),$$

where $\mu = k_3 Y_{\text{tot}}^2/k_2, \lambda = k_1/k_2$. Determine the existence and stability of the fixed points for y. Plot the bifurcation diagram with μ treated as a bifurcation parameter and $\lambda = 0.03$. Hence, show that the system is bistable over a range of values of μ.

(b) Suppose that there are N molecules, that is, $N = VY_{\text{tot}}$, where V is cell volume. Construct the master equation for the probability $P(n,t)$ that there are $n(t) = n$ molecules in state Y at time t.

(c) Using equation (3.2.6), show that the steady-state distribution is

$$P_s(n) = \frac{C_N N!}{n!(N-n)!} \prod_{m=0}^{n-1} \left[\lambda + \frac{\mu}{N^2} m(m-1)\right].$$

Plot $P_s(n)$ as a function of n (treated as a continuous variable over the range $[0,400]$) for $N = 400$, $\mu = 4.5$, and $\mu = 6$ with $\lambda = 0.03$. Comment on the location of the peaks in terms of fixed points of the deterministic system.

(d) Derive the corresponding FPE using a Kramers–Moyal expansion and determine the steady-state solution. Calculate the steady-state solution and compare with the exact solution of part (c) for $N = 40$ and $N = 400$.

Problem 3.8 (Mean time to escape). Consider the model of bistability in a population of two-state ion channels analyzed in Sect. 3.3 (see also Fig. 3.4). The FPT to escape from the state x_- is

$$\tau(x_-) = N \int_{x_-}^{x_0} \mathrm{e}^{N\Psi(x')} dx' \int_0^{x'} \frac{\mathrm{e}^{-N\Psi(x'')}}{B(x'')} dx'',$$

with the quasi-potential

$$\Psi(x) = \int_0^x \frac{A(x')}{B(x')} dx'.$$

By following the steps in the derivation of the Arrhenius formula (3.3.32), show that the rate of escape is given by

$$\lambda = \frac{1}{\tau(x_-)} \approx \frac{B(x_-)}{2\pi} \sqrt{\left|\frac{A'(x_-)}{B(x_-)}\right| \left|\frac{A'(x_0)}{B(x_0)}\right|} \exp\left[N \int_{x_-}^{x_0} \frac{A(x)}{B(x)} dx.\right].$$

Problem 3.9 (De Young–Keizer model). Carry out the reduction of the De Young–Keizer model discussed in Sect. 3.4.1.

(a) First write down the kinetic equations for the four states without calcium bound to the inactivating site $(S_{000}, S_{010}, S_{100}, S_{110})$, using the reaction diagram of Fig. 3.6a.

(b) Perform the QSS approximation by setting all time derivatives to zero and dropping all slow transitions involving binding/unbinding of the calcium inactivating binding site, that is, set $k_{\pm 2} = k_{\pm 4} = 0$. Show that

$$x_{100} = \frac{K_5 h}{c + K_5} - x_{000}, \quad x_{010} = \frac{K_1 h}{p + K_1} - x_{000},$$

where $K_i = k_{-i}/k_i$ and

$$x_{000} + x_{010} + x_{100} + x_{110} = h.$$

Hence, determine x_{000} and x_{110}.

(c) Show that the corresponding QSS solutions for the states with calcium bound to the inactivating site $(S_{001}, S_{011}, S_{101}, S_{111})$ are obtained from (b) by taking $K_1 \to K_3$ and $h \to 1 - h$. Note that

$$x_{001} + x_{011} + x_{101} + x_{111} = 1 - h.$$

(d) Finally, add together the four kinetic equations of part (a) and substitute for the x_{ijk} using their QSS solutions. Hence derive the equation

$$\frac{dh}{dt} = -\frac{(k_{-4}K_2K_1 + k_{-2}pK_4)c}{K_4K_2(p+K_1)}h + \frac{k_{-2}p + k_{-4}K_3}{p+K_3}(1-h).$$

Problem 3.10 (Method of characteristics). Recall that the characteristic function for the Ornstein–Uhlenbeck process satisfies the linear PDE (see Ex. 2.5)

$$\frac{\partial \Gamma}{\partial t} + kz\frac{\partial \Gamma}{\partial z} = -\frac{D}{2}z^2\Gamma.$$

Use the method of characteristics to obtain a solution of the form

$$\Gamma(z,t) = \Gamma_0(z\mathrm{e}^{-kt})\mathrm{e}^{-Dz^2/4k}$$

and show how Γ_0 is related to the initial condition for p.

Problem 3.11 (Gating model of confinement). Consider the gating model of confinement within a corral that was introduced in Sect. 3.6. If the gate is always open, then the generating function for stochastic confinement is given by

$$G(z,t) = [1+\mathrm{e}^{-\gamma t}(z-1)]^{n_0}\mathrm{e}^{k\left(1-\mathrm{e}^{-\gamma t}\right)(z-1)/\gamma}.$$

(a) Suppose that $n_0 = 0$ (domain is initially empty). Show that $P_n(t)$ is given by a time-dependent Poisson process with rate

$$\lambda(t) = \frac{k}{\gamma}(1-\mathrm{e}^{-\gamma t})$$

by Taylor expanding $G(z,t)$ as a function of z.

(b) Now suppose that $n_0 \neq 0$. Determine the mean and variance using the formulae

$$\langle n(t)\rangle = \left.\frac{\partial G(z,t)}{\partial z}\right|_{z=1}, \quad \langle (n^2(t) - n(t))\rangle = \left.\frac{\partial^2 G(z,t)}{\partial z^2}\right|_{z=1}.$$

Problem 3.12 (Computer simulations: two-state ion channels). In this problem we investigate the diffusion approximation of the master equation (3.2.1) for an ensemble of two-state ion channels:

$$\frac{d}{dt}P(n,t) = \alpha(N-n+1)P(n-1,t) + \beta(n+1)P(n+1,t) - [\alpha(N-n)+\beta n]P(n,t).$$

Take $\alpha = 1, \beta = 2$, and $N = 100$.

(a) Numerically solve the master equation using Euler's direct method for $t \in [0,1]$ and $\Delta t = 0.01$. Plot the histogram of $P_n(T)$ for $T = 1$ and compare with the steady-state distribution (3.2.8).
(b) Use the Gillespie algorithm of Sect. 6.8 to generate sample paths for the number $n(t)$ of open ion channels for $t \in [0,10]$. The two reactions are $n \to n+1$ at a rate $\alpha(N-n)$ and $n \to n-1$ at a rate βn. By averaging over sample paths, compare the histogram of $n(T)$ with the distribution $P_n(T)$ for $T = 1$.
(c) Use Euler's direct method (see Sect. 2.6.6) to simulate the Langevin equation

$$dX(t) = [\alpha(1-X) - \beta X]dt + \frac{1}{\sqrt{N}}\sqrt{\alpha(1-X)+\beta X}dW(t),$$

obtained by carrying out a Kramers–Moyal expansion of the master equation. Here $X(t)$ is the fraction of open ion channels at time t. Construct a histogram of $X(T)$ for $T = 1$ and compare with the results of part (b). Repeat for $N = 10$ and $N = 1,000$ and comment on the differences.

Problem 3.13 (Computer simulations: noise-induced escape). Consider the simplified stochastic conductance-based model given by equation (3.5.15):

$$C_m \frac{dV}{dt} = \frac{n}{N} g_{\rm Na}[V_{\rm Na} - V] + g_{\rm eff}[V_{\rm eff} - V] + I.$$

The transition rates for the opening and closing of the Na channels are

$$\omega_+ = \beta \exp[(v+1.2)/9](N-n), \quad \omega_- = \beta n.$$

Take the parameter values $N = 10$, $C_m = 20\,\mu$F /cm^2, $V_{\rm Na} = 120$ mV, $V_{\rm eff} = -62.3$ mV, $g_{\rm Na} = 4.4$ mS/cm^2, $g_{\rm eff} = 2.2$ mS/cm^2, and $\beta = 0.8$ s^{-1}.

(a) Adapt the Gillespie algorithm (Sect. 6.8) to simulate sample trajectories of the above stochastic hybrid system. That is, use Gillespie to determine the random sequence of times at which a jump occurs and whether the number of channels increases or decreases by unity at each jump. In between jumps, solve the equation for V given the current state n.
(b) Pick a value of current I (in mA) for which the deterministic system is bistable. Taking the initial condition to be the stable rest state, determine the first time for the system to reach the unstable fixed point. Repeat over many trials in order to estimate the MFPT of escape. Compare with the Arrhenius formula (3.3.34) for $N = 10$ and $N = 100$.

Chapter 4
Polymers and Molecular Motors

The *cytoskeleton* within the cytoplasm plays important roles in maintaining the structural integrity of a cell, intracellular transport, cell motility, and cell division. In eukaryotes, the cytoskeleton consists of three types of protein filaments—microtubules, intermediate filaments, and actin filaments (see Fig. 4.1). *Actin filaments* are the thinnest structures (around 6 nm) whose basic building block is the globular protein G-actin. These can assemble into a long filamentous chain known as F-actin, which has the superficial appearance of two interlocked strands. Actin filaments are relatively flexible and strong. (The degree of flexibility of a polymer can be characterized in terms of the so-called persistence length ξ_p, which characterizes the length scale over which correlations in the orientation of the polymer persist. If $\xi_p \gg L$, where L is the total length of the polymer, then the polymer is said to be rigid, whereas if $\xi_p \ll L$, then it acts like a random coil.) Actin dynamics plays a major role in cell motility, where one end (the $+$ or barbed end) elongates due to polymerization while the other end (the $-$ or pointed end) contracts due to a combination of depolymerization and myosin motors (Sect. 8.3). F-actin also serves as a tensile platform for myosin motors involved in the pulling action of muscle contraction. Actin filaments are themselves assembled into two general types of structures: bundles called *filopodia* that consist of parallel arrays of filaments and cross-linked networks called *lamellipodia*. *Microtubules* are hollow cylinders around 23 nm in diameter, which typically consist of 13 protofilaments, each of which is a polymer made up of heterodimers of alpha and beta tubulin. Microtubules project radially from organizing centers known as centrosomes and play a key role in cell division via the mitotic spindle (Sect. 8.2). Finally, *intermediate filaments* average 10 nm in diameter, are more strongly bound than F-actin, and act to maintain the structural integrity of a cell. Actin and tubulin filaments are assembled via the polymerization of subunits, which change their chemical state when incorporated into a filament. For example, actin monomers contain an ATP molecule that rapidly hydrolyzes to adenosine diphosphate (ADP) following polymerization. Similarly, the β unit of the tubulin heterodimer contains a guanosine triphosphate (GTP) molecule that hydrolyzes to guanosine diphosphate (GDP) after polymerization. These chemical

P.C. Bressloff, *Stochastic Processes in Cell Biology*, Interdisciplinary Applied Mathematics 41, DOI 10.1007/978-3-319-08488-6_4

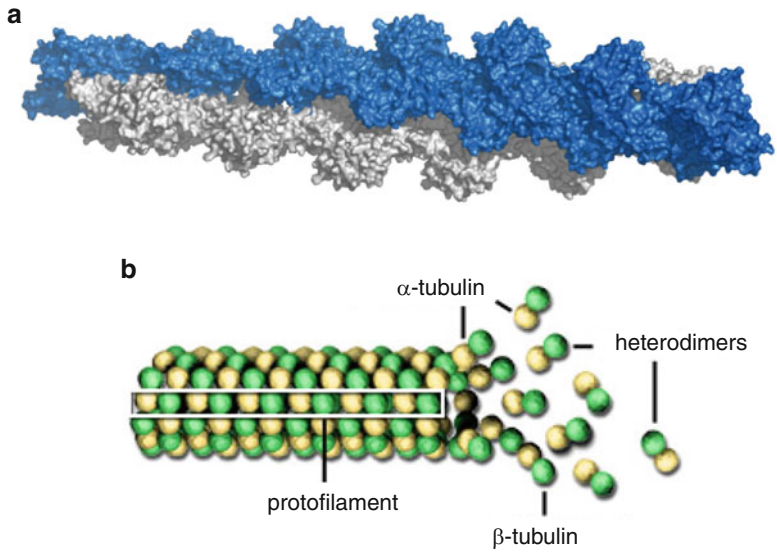


Fig. 4.1: Cell cytoskeletal filaments. (**a**) Computer reconstruction of the atomic structure of an actin filament with 13 subunits by Thomas Splettstoesser using open source software PyMol. (**b**) Schematic illustration of helical structure of a microtubule. Public domain figures downloaded from Wikipedia Commons

transformations can lead to more complex phenomena than observed in simple polymers, such as treadmilling and dynamical instabilities.

Another major function of actin and microtubular polymer filaments is that they act as effective 1D tracks for the active movement of *molecular motor proteins*. Diffusion inside the cytosol or within the plasma membrane of a cell is a means by which dissolved macromolecules can be passively transported without any input of energy. However, there are two main limitations of passive diffusion as a mechanism for intracellular transport:

(i) It can take far too long to travel the long distances necessary to reach targets within a cell, which is particularly acute in the case of the axons and dendrites of neurons.
(ii) Diffusive transport tends to be unbiased, making it difficult to target resources to specific areas within a cell.

Active intracellular transport can overcome these difficulties so that movement is both faster and direction specific, but does so at a price. Active transport cannot occur under thermodynamic equilibrium, which means that energy must be consumed by this process, typically via the hydrolysis of ATP. The main types of active intracellular transport involve the molecular motors *kinesin* and *dynein* carrying resources along microtubular filament tracks and *myosin V* motors transporting cargo along actin filaments. As we have already highlighted, microtubules and actin

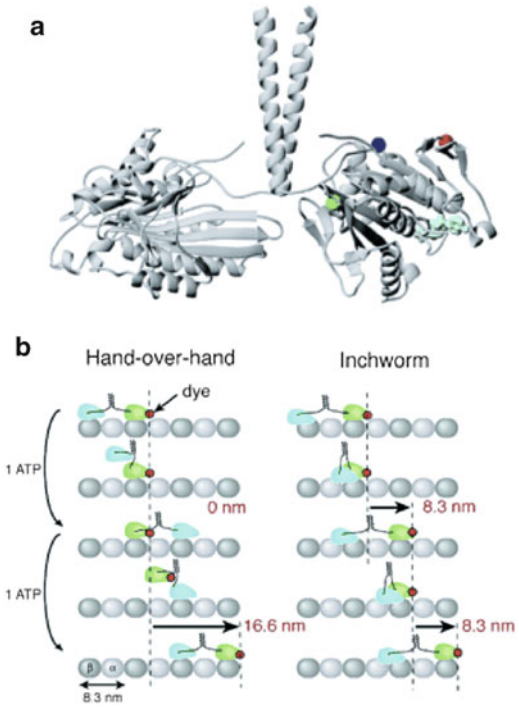


Fig. 4.2: (**a**) The kinesin molecule consists of two motor units (analogous to feet) that are linked together via a common stalk that attaches to cargo to be transported. (**b**) Previous studies had shown that the kinesin molecule moves along the microtubule in steps of 8 nm. Kinetic analysis of the dwell time between steps shows that there is an alternation of displacement from one step to the other, supporting a hand-over-hand model rather than an inchworm model. In the latter case, both feet would move only in 8-nm steps as the kinesin molecule's center of mass moves. However, if the kinesin molecule moves in a hand-over-hand motion, then the "rear" foot should take a 16-nm step forward during one cycle and then 0 nm during the next cycle. (Adapted from Yildiz et al. [695].)

filaments are polarized polymers with biophysically distinct $(+)$ and $(-)$ ends. It turns out that this polarity determines the preferred direction in which an individual molecular motor moves. For example, kinesin moves towards the $(+)$ end whereas dynein moves towards the $(-)$ end of a microtubule. Each motor protein undergoes a sequence of conformational changes after reacting with one or more ATP molecules, causing it to step forward along a filament in its preferred direction (see Fig. 4.2). Thus, ATP provides the energy necessary for the molecular motor to do work in the form of pulling its cargo along a filament in a biased direction. When modeling active transport, one usually neglects the dynamics of microtubules and actin filaments and simply treats them as static 1D tracks with periodic structure. On the other hand, the regulation of polymerization and depolymerization by molecular motors plays an important role in the formation and maintenance of certain cytoskeletal structures such as the mitotic spindle (Chap. 8).

The movement of molecular motors such as kinesin occurs over several length and time scales [309, 325, 348, 391]. In the case of a single motor there are at least three regimes:

(a) The mechanochemical energy transduction process that generates a single step of the motor. In the case of dimeric or double-headed kinesin, a single step is of length 8 nm and the total conformational cycle takes around 10 ms.
(b) The effective biased random walk along a filament during a single run, in which the motor takes multiple steps before dissociating from the filament. For example, kinesin takes around 100 steps in a single run, covering a distance of around 1 μm. Walking distances can be substantially increased if several molecular motors pull the cargo.
(c) The alternating periods of directed motion along the filament and diffusive or stationary motion when the motor is unbound from the filament. In the unbound state a motor diffuses in the surrounding aqueous solution with a diffusion coefficient of the order $1\,\mu m^2\,s^{-1}$. However, molecular crowding tends to confine the motor so that it stays close to its detachment point. In the case of multiple molecular motors transporting cargo, the resulting complex can exhibit bidirectional motion [244, 359, 577, 677].

Advances in experimental techniques have generated considerable information about the structural properties of molecular motors and their dynamics. For example, optical traps (Sect. 1.2) have been used to measure how changes in ATP concentration affect the force–displacement properties of both kinesin [571, 658] and dynein [201, 335]. A sketch of typical results obtained for kinesin is shown in Fig. 4.3. Such data can be incorporated into models at levels (b) and (c). On the other hand, information about the energetics of the various conformational states and the rates of transitions between them are not yet sufficient to develop detailed biophysical models of motors. Hence, it is not possible to generate realistic velocity–force curves, for example, without considerable data fitting. Thus much of the work on molecular motors at the smallest scale (a) is of a more qualitative nature, in which one tries to understand the basic principles that allow nanoscale machines to do useful work in the presence of thermal noise—so-called Brownian ratchet models.

In addition to intracellular transport, molecular motors perform many other functions within a cell:

(i) Muscle contraction and cell locomotion due to the collective action of multiple myosin II motor heads (cross bridges) interacting with actin filaments (see Chap. 15 of Keener and Sneyd [322])
(ii) The reversible action of rotary motor ATP synthase, which either produces ATP using ion gradients or acts as an ion pump fueled by ATP hydrolysis [166]
(iii) The swimming and tumbling of bacteria such as *E. coli* driven by flagella rotary motors (see Sect. 5.3)

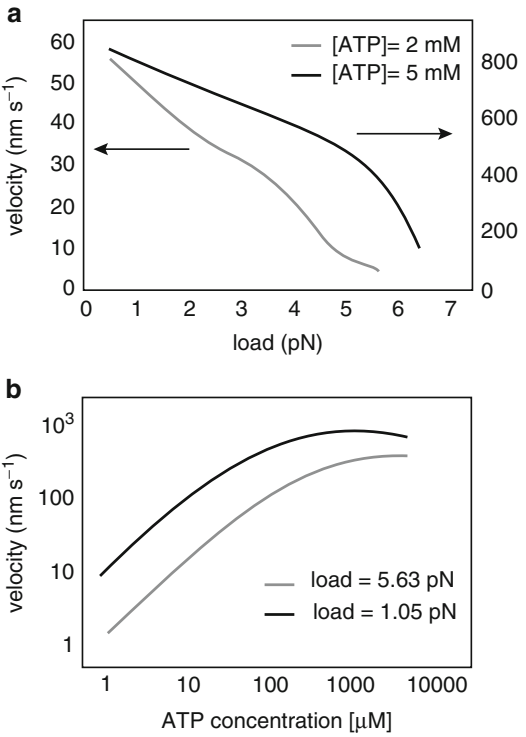


Fig. 4.3: Sketch of results from optical trap measurements of kinesin. (**a**) Variation of velocity with load for high and low ATP concentrations. (**b**) Variation of velocity with ATP concentration for low and high loads. (Redrawn from Visscher et al. [658].)

(iv) Transcription of RNA from DNA via RNAP [669] (Sect. 6.6)
(v) The action of viral DNA packaging motors that inject viral genomic DNA into the protein shell (capsid) of a bacteriophage (a virus that infects and replicates within bacteria) as part of its replication cycle [697]

In this chapter we introduce some basic stochastic models of polymerization and of molecular motors, involving both discrete and continuous Markov processes. The dynamics of polymers will figure significantly in Chap. 8 on the self-organization of the cytoskeleton, whereas the role of molecular motors in active intracellular transport will be considered in Chap. 7. For a comprehensive introduction to the mechanics of motor proteins and the cytoskeleton, see the book by Howard [275]. We begin by considering a simple 1D stochastic model of polymerization and depolymerization, which neglects molecular details such as the structure of heterodimers and helical protofilaments (Sect. 4.1). The model takes the form of a birth–death process that keeps track of the addition or removal of monomers from one or both

ends of the polymer. We also consider two models of microtubule catastrophe, a two-state model of growth and shrinking phases and a more detailed model of the effects of ATP hydrolysis and cap formation.

In Sect. 4.2 we introduce the theory of Brownian ratchets by considering the FPE for a Brownian particle moving in a periodic ratchet (asymmetric) potential. We show that the mean velocity of the Brownian particle is zero, which implies that such a potential cannot provide a mechanism for a molecular motor to do useful work against an applied load. One mechanism for breaking the periodicity is to rectify the motion, as exemplified by polymerization ratchets and translocation ratchets; energy is provided by the binding of a molecule to the polymer or protein. Polymerization ratchets play a major role in cell motility (Sect. 8.4) and cell polarization (Sect. 9.5), where the force exerted by an actin filament extrudes the cell membrane in a particular direction. On the other hand, the translocation ratchet is used to model the transport of a polymer through a membrane pore. On one side of the membrane, proteins known as chaperones, which are too large to pass through the pore, bind the polymer and thus rectify its motion through the pore (see also Sect. 7.3).

In Sect. 4.3 we describe a qualitative model of processive molecular motors such as kinesin and dynein that is based on a two-state Brownian ratchet (flashing ratchet). The basic idea is that the motor has to negotiate a periodic potential energy landscape based on its interactions with the microtubule filaments and the form of the landscape depends on the conformational state of the motor. (The idea of representing a molecular motor in terms of several conformational states that depend on interactions with a filament was first introduced by Huxley in his theoretical study of muscles [290].) We show that useful work can be generated provided that the transition rates between the different conformational states do not satisfy detailed balance, which is achieved via the hydrolysis of ATP. We end by briefly describing an alternative, kinetic approach to modeling the stepping of molecular motors, based on a discrete Markov process. The state transition diagram includes both jumps between conformational states and jumps between neighboring sites on the filament.

We further develop the theory of molecular motors in Sect. 4.4, where we consider three examples of the collective motion of an ensemble of molecular motors: (i) the tug-of-war (ToW) model of bidirectional vesicular transport by opposing groups of processive motors; (ii) the motor-assisted extraction of membrane nanotubes; (iii) a model of interacting motors attached to a rigid cytoskeletal backbone. We show how the last two models support collective oscillations consistent with those seen experimentally. We also review the Hopf bifurcation theorem and linear stability analysis, since the Hopf bifurcation is the underlying mechanism for the emergence of these collective oscillations. Finally, in the appendix (Sect. 4.5) we provide a brief introduction to the statistical mechanics of polymers, building upon the theory briefly summarized in Sect. 1.4. We focus on how entropic forces arise from the uncoiling of a polymer, since it leads to a reduction in the number of possible configurational states. Some of the ideas will later be used to study the effects of entropic forces on polymer translocation through a membrane pore and diffusion through narrow channels (Sect. 7.3).

4.1 Polymerization

4.1.1 Simple Model of a Single Polymer Filament

Consider, for simplicity, monomers binding or unbinding at the + end of a single-stranded filament (see Fig. 4.4 and Ex. 4.1). (For an extension to multi-stranded filaments see [614] and Ex. 4.2.) Suppose that the minimum length of the polymer is either a single monomer or a critical nucleus of M monomers, which for the moment is considered stable. Let n, $n \geq 0$, denote the number of monomers added to this critical nucleus and take the rate of monomer binding and unbinding to be π and ε, respectively. The probability $P_n(t)$ that the filament contains n additional monomers at time t satisfies the master equation

$$\frac{dP_n}{dt} = \varepsilon P_{n+1}(t) + \pi P_{n-1}(t) - [\varepsilon + \pi]P_n(t), \quad n > 0 \tag{4.1.1}$$

supplemented by the reflecting boundary condition

$$\frac{dP_0}{dt} = \varepsilon P_1(t) - \pi P_0(t), \tag{4.1.2}$$

and the normalization condition $\sum_{n=0}^{\infty} P_n(t) = 1$. We are assuming that there is an unlimited supply of monomers. First, note that if we multiply both sides of Eq. (4.1.1) by n and sum over n, then we obtain a mean-field equation for the mean change in length $\langle n \rangle = \sum_{n=0}^{\infty} nP_n$, namely,

$$\frac{d\langle n \rangle}{dt} = \pi - \varepsilon + \varepsilon P_0. \tag{4.1.3}$$

If the master equation has a stationary solution, then

$$\varepsilon P_{n+1} + \pi P_{n-1} - [\varepsilon + \pi]P_n = 0, \quad n > 0, \quad \varepsilon P_1 = \pi P_0.$$

Since the binding/unbinding rates are n-independent, the solution is of the form $P_n = C\lambda^n$. Substituting this into the stationary equation for $n > 0$ gives

$$\varepsilon \lambda^{n+1} + \pi \lambda^{n-1} - [\varepsilon + \pi]\lambda^n = 0, \quad n > 0,$$

which reduces to the quadratic equation

$$\lambda^2 + r\lambda - (1 + r) = 0, \quad r = \frac{\pi}{\varepsilon}.$$

This has the solutions $\lambda = 1$ or $\lambda = r$. The normalization condition $C\sum_{n=0}^{\infty} \lambda^n = 1$ requires that $\lambda < 1$ so that there exists a unique stationary solution provided that $\pi/\varepsilon < 1$. Solving for C using the normalization condition then gives

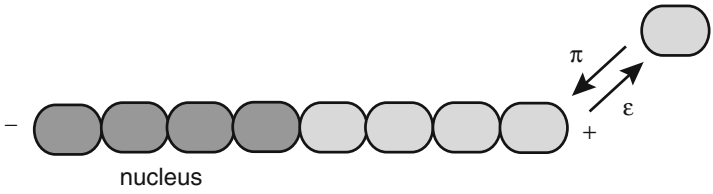


Fig. 4.4: Simple model of F-actin undergoing polymerization at one end

$$P_n = \left(1 - \frac{\pi}{\varepsilon}\right)\left(\frac{\pi}{\varepsilon}\right)^n, \quad n \geq 0. \tag{4.1.4}$$

It also follows that $d\langle n\rangle/dt = 0$.

So far we have considered a stochastic model of a single filament. Each model simulation generates a sample path of the stochastic behavior, and statistics can be extracted by running many trials. An alternative picture is to consider a large population of N identical filaments. Suppose that each filament is in the same initial state. For sufficiently large N, we expect the number $X_n(t)$ of filaments having additional length n to be $X_n(t) = NP_n(t)$. Since the transition rates are n-independent, we can simply multiply the master equation by N to obtain corresponding kinetic equations for the X_n with $\sum_{n=0}^{\infty} X_n(t) = N$. As we have discussed in the context of chemical reactions, the kinetic equations for a population of filaments are deterministic. If one wanted to take into account fluctuations due to intrinsic noise, then one would have to consider the master equation for the probability distribution $\mathbb{P}(M_0, M_1, \ldots, M_N, t)$ where M_n is the number of filaments in state n.

The deterministic population model has been extended to take into account the disappearance and production of critical nuclei of size M [156, 282]. Taking X_n to denote the fraction of filaments of length n, the kinetic equations are

$$\frac{dX_n}{dt} = \varepsilon X_{n+1}(t) + \pi X_{n-1}(t) - [\varepsilon + \pi]X_n(t), \quad n > M \tag{4.1.5}$$

and

$$\frac{dX_M}{dt} = \varepsilon X_{M+1}(t) - (\pi + \varepsilon)X_M(t) + \sigma, \tag{4.1.6}$$

where nuclei can disappear (convert back to M monomers) at a rate ε and are produced at a rate σ. Assuming a fixed background monomer concentration a, the binding and production rates are taken to be

$$\pi = \pi_0 a, \quad \sigma = \sigma_0 a^M, \tag{4.1.7}$$

with σ_0, π_0 independent of a. One no longer has a conservation condition for the total number of filaments. However, if $\pi/\varepsilon < 1$, then one can still construct a steady-state solution of the form $X_n = C(\pi/\varepsilon)^n$. Substituting into the steady-state solution for $n = M$, we have

$$C\varepsilon\left(\frac{\pi}{\varepsilon}\right)^{M+1} - C(\varepsilon + \pi)\left(\frac{\pi}{\varepsilon}\right)^M + \sigma = 0,$$

which implies that $C = (\sigma/\varepsilon)(\varepsilon/\pi)^M$ and

$$X_n = \frac{\sigma}{\varepsilon}\left(\frac{\pi}{\varepsilon}\right)^{n-M} = \frac{\sigma_0}{\varepsilon}\left(\frac{\pi_0}{\varepsilon}\right)^{n-M} a^n. \tag{4.1.8}$$

It immediately follows that the mean filament length L in the population is

$$L = M + \frac{\sum_{n=M}^{\infty}(n-M)X_n}{\sum_{n=M}^{\infty} X_n} = M + \frac{\sum_{n=0}^{\infty} n(\pi/\varepsilon)^n}{\sum_{n=0}^{\infty}(\pi/\varepsilon)^n} = M + \frac{\pi/\varepsilon}{1-\pi/\varepsilon}.$$

Hence, the mean length diverges as $\pi \to \varepsilon$ from below.

It is also possible to analyze the stability of the steady state in the case that the polymers have a maximum size $n = J$. After dividing the kinetic equation by ε and rescaling time, we obtain the matrix equation

$$\frac{d\mathbf{X}}{dt} = \mathbf{M}\mathbf{X} + \mathbf{s},$$

where $\mathbf{X} = (X_M, X_{M+1}, \ldots, X_J)^T$, $\mathbf{s} = (\sigma/\varepsilon, 0, \ldots, 0)^T$ and $\mathbf{M}$ is the tridiagonal matrix

$$\mathbf{M} = \begin{pmatrix} -(1+r) & 1 & 0 & 0 \cdots 0 \\ r & -(1+r) & 1 & 0 \cdots 0 \\ 0 & r & -(1+r) & 1 \cdots 0 \\ \vdots & \vdots & \vdots & \vdots\ \vdots\ \vdots \\ 0 & \cdots & \cdots & 0\ \ r\ \ 1 \end{pmatrix}.$$

This linear system has a general solution of the form

$$\mathbf{X}(t) = \mathbf{X}_0 + \sum_{j=1}^{J-M+1} \mathbf{v}_j e^{\lambda_j t},$$

where $\mathbf{X}_0$ is the steady-state solution and $0, \lambda_j$, $j = 1, J+1-M$ are the eigenvalues of $\mathbf{M}$. We can now use same basic results from linear algebra. First, since $\mathbf{M}$ is tridiagonal with $M_{m,n+1}M_{n+1,n} = r > 0$, it follows that the eigenvalues are real and simple. In particular, none of the eigenvalues λ_j vanish. The *Gershgorin disk theorem* can then be used to establish that none of the eigenvalues λ_j are positive definite and are thus negative definite. The theorem states that the eigenvalues of the tridiagonal matrix $\mathbf{M}$ are contained in the union of disks D_n in the complex λ-plane with

$$D_n = \{|\lambda - M_{nn}| \le \sum_{k \ne n} |M_{nk}|\}.$$

The first disk D_1 has center at $\lambda = -(1+r)$ and radius 1, whereas all the disks $D_2, \ldots, D_{J-1}$ have centers at $\lambda = -(1+r)$ and radii $1+r$. All of these lie in the

left-half complex plane. Finally, D_J has a center at $\lambda = -1$ and radius r, which also lies the left-half complex plane provided that $r \leq 1$.

Now suppose that actin monomers can bind or unbind at both ends with rates $k_{\text{on}}^{\pm}$ and $k_{\text{off}}^{\pm}$, as shown in Fig. 4.5. The binding rate is multiplied by a fixed background monomer concentration a. (The spatial effects of a nonuniform monomer concentration are considered by Edelstein-Keshet and Ermentrout [157]; see also Ex. 4.3.) The difference between the two ends is due to the fact the ATP-actin quickly hydrolyzes to ATD-actin so that the tip consists of ATP-actin and the tail consists of ATD-actin. Rather than writing down the master equation for the system, let us consider the equations for the mean number of monomers $n_{\pm}$ added at each end. Assuming that the filament is sufficiently long, we have

$$\frac{dn_+}{dt} = k_{\text{on}}^{+} a - k_{\text{off}}^{+}, \quad \frac{dn_-}{dt} = k_{\text{on}}^{-} a - k_{\text{off}}^{-}. \tag{4.1.9}$$

It is clear that the $\pm$ end grows provided that $a > a_c^{\pm}$, where $a_c^{\pm} = k_{\text{off}}^{-}/k_{\text{on}}^{-}$. If $a_c^{+} \approx a_c^{-}$, then both ends shrink or grow simultaneously. On the other hand, if $a_c^{+} < a < a_c^{-}$ then the plus end grows at the same time the minus end shrinks. Finally, adding the pair of Eq. (4.1.9) shows that

$$\frac{dn}{dt} = k_{\text{on}} a - k_{\text{off}},$$

with $n = n_+ + n_-$, $k_{\text{off}} = k_{\text{off}}^{+} + k_{\text{off}}^{-}$, and $k_{\text{on}} = k_{\text{on}}^{+} + k_{\text{on}}^{-}$. Hence, if the monomer concentration $a = a_0$, where

$$a_0 = \frac{k_{\text{off}}^{+} + k_{\text{off}}^{-}}{k_{\text{on}}^{+} + k_{\text{on}}^{-}} \equiv \frac{k_{\text{off}}}{k_{\text{on}}},$$

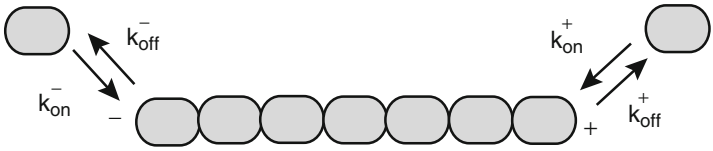


Fig. 4.5: Model of F-actin undergoing polymerization at both ends

then the total filament length remains constant even though monomers are constantly moving along its length—treadmilling.

4.1.2 Microtubule Catastrophes

An interesting aspect of microtubules is that they undergo periods of persistent microtubule growth interrupted by occasional switching to rapid shrinkage know as "microtubule catastrophe" [443]. Microtubules grow by the attachment of GTP-tubulin complexes at one end. In order to maintain growth, the end of the microtubule must consist of a "cap" of consecutive GTP-tubulin monomers. However, each polymerized complex can hydrolyze into GDP-tubulin such that if all the monomers in the cap convert to GDP, then the microtubule is destabilized, and there is rapid shrinkage due to detachment of the GTP-tubulin monomers. The competition between attachment of GTP-tubulin and hydrolysis from GTP to GTD is thought to be the basic mechanism of alternating periods of growth and shrinkage.

Two-State Model

One approach to modeling catastrophe is based on an effective two-state model, in which the microtubule exists either in a growing phase or a shrinking phase and

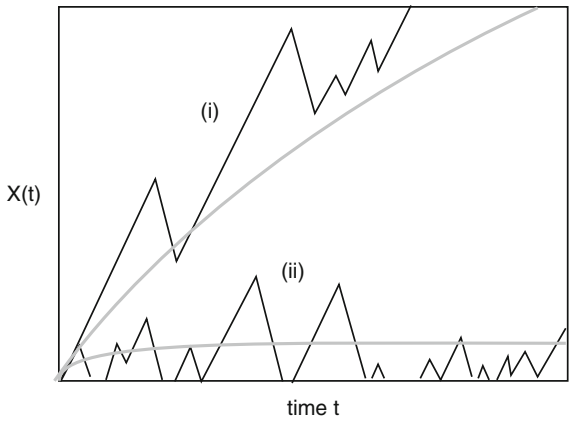


Fig. 4.6: Schematic diagram illustrating two phases of microtubule growth depending on the sign of the mean growth rate V. (*i*) For $V > 0$ the microtubule undergoes unbounded growth. (*ii*) For $V < 0$ the frequency of catastrophes increases so that there is bounded growth. The mean length as a function of time is shown by *gray curves*

there are stochastic transitions between the two states based on a Markov process [50, 146]. The microscopic details of cap formation and hydrolysis are not modeled explicitly. It is assumed that one end of a microtubule is fixed and the position of the other end is taken to be a stochastic variable $X(t)$, which can also be identified as the variable length of the microtubule. Let $P_{\pm}(x,t)$ be the probability density that at time t the end of the microtubule is at $X(t) = x$ and it is in either the growing

phase (+) or the shrinking phase (−). The Dogterom–Leibler model of microtubule catastrophe takes the form [146]

$$\frac{\partial P_+}{\partial t} = -v_+ \frac{\partial P_+}{\partial x} - k_+ P_+ + k_- P_- \tag{4.1.10a}$$

$$\frac{\partial P_-}{\partial t} = v_- \frac{\partial P_-}{\partial x} - k_- P_- + k_+ P_+. \tag{4.1.10b}$$

Here v_+ and v_- are the average speeds of growth and shrinking and $k_\pm$ are the transition rates between the two states. Both v_+ and k_- are assumed to vary linearly with the tubulin concentration c at the tip of the microtubule, $k- = kc, v_+ = cu$; the concentration c is taken to be fixed. We can determine a condition for the existence of a steady state by adding Eqs. (4.1.10a) and (4.1.10b) and setting $\partial_t P_\pm = 0$. This shows that

$$v \frac{\partial P_+}{\partial x} - v \frac{\partial P_-}{\partial x} = 0,$$

and thus $v_+ P_+ - v_- P_- =$ constant. Assuming a semi-infinite domain for x, normalizability of $P_\pm$ implies that the constant must be zero and, hence, $P_\pm = P/v_\pm$ with P satisfying the equation

$$\frac{dP(x)}{dx} = \left[\frac{k_-}{v_-} - \frac{k_+}{v_+}\right] P(x) = \frac{V}{D} P(x),$$

where

$$V = \frac{k_- v_+ - k_+ v_-}{k_+ + k_-}$$

is the mean speed of microtubular growth (based on the steady-state solution of the two-state Markov process describing the switching between growth and shrinking phases) and

$$D = \frac{v_+ v_-}{k_+ + k_-}$$

is an effective diffusivity. It immediately follows that there exists a steady-state solution, $P(x) = P(0)\mathrm{e}^{-Vx/D}$, $0 < x < \infty$, if and only if $V < 0$. In the regime $V > 0$, catastrophe events are relatively rare and the microtubule continuously grows with mean speed V, whereas, for $V < 0$, the catastrophe events occur much more frequently so that there is a balance between growth and shrinkage that results in a steady-state distribution of microtubule lengths (see Fig. 4.6).

Model of Hydrolysis and Caps

Another approach to modeling microtubule dynamics is to include a simplified description of hydrolysis and cap formation that involves only a few model parameters [11, 185, 186]. Here we will describe in some detail the model of Antal et al. [11].

The microtubule is taken to consist of a mixture of GTP-tubulin complexes (GTP-T) and GTD-tubulin complexes (GTD-T). A given configuration is represented by a string of $+$ and $-$ symbols corresponding to GTP-T and GTD-T, respectively. Three basic processes are considered:

1. Attachment: Growth of a microtubule occurs via the attachment of a GTP-T monomer at one end, with the attachment rate depending on the identity of the current monomer at the tip. That is,

$$|\ldots+\rangle \Longrightarrow |\ldots++\rangle \text{ at rate } \lambda, \quad |\ldots-\rangle \Longrightarrow |\ldots-+\rangle \text{ at rate } p\lambda,$$

   with $p \leq 1$.

2. Conversion: Once incorporated into the microtubule, each GTP-T can independently convert by hydrolysis to GTD-T:

$$|\ldots+\ldots\rangle \Longrightarrow |\ldots-\ldots\rangle \text{ at rate } 1.$$

3. Detachment: Shrinkage of a microtubule occurs via the detachment of a GTD-T monomer from the end of the microtubule

$$|\ldots-\rangle \Longrightarrow |\ldots\rangle \text{ at rate } \mu.$$

In general, one finds that there are two phases in the parameter space (λ, μ, p), one corresponding to a growing phase with average growth rate $V(\lambda, \mu, p)$ and the other to a bounded phase. The two phases are separated by a phase boundary $\mu = \mu_*(\lambda, p)$ along which $V = 0$. Following Antal et al. [11], we will develop the stochastic analysis of the model by considering various limiting cases.

Unconstrained Growth. First, suppose that there is unrestricted growth ($\mu = 0$) and the attachment rate is independent of the end monomer ($p = 1$). The speed of growth is then simply $V = \lambda$. The kinetic equation for the mean number x of GTP-T monomers in the chain is

$$\frac{dx}{dt} = \lambda - x.$$

The probability $p_m(t)$ that there are m GTP-T monomers at time t evolves according to a birth–death master equation (6.2.1):

$$\frac{dp_m}{dt} = (m+1)p_{m+1}(t) + \lambda p_{m-1}(t) - (\lambda + m)p_m(t).$$

This is identical in form to Eq. (3.6.3), which means that $p_m(t)$ is given by the time-dependent Poisson distribution

$$p_m(t) = \frac{\lambda(t)^m \mathrm{e}^{-\lambda(t)}}{m!}, \quad \lambda(t) = \lambda(1 - \mathrm{e}^{-t}), \tag{4.1.11}$$

Fig. 4.7: Schematic diagram illustrating a cap of GTP-T monomers and additional islands of GTP-T

and

$$\langle m(t)\rangle = \mathrm{var}[m(t)] = \lambda(1-\mathrm{e}^{-t}).$$

Cap Length. The conversion of GTP-T to GTD-T means that more recently attached monomers around the tip region are more likely to be GTP-T, whereas monomers in the tail are predominantly GTD-T. The cap is defined to be the region from the end of the microtubule to the first GTD-T monomer (see Fig. 4.7). Let π_k be the probability that the cap is of length k and consider the associated master equation

$$\frac{d\pi_k}{dt} = \lambda(\pi_{k-1}-\pi_k) - k\pi_k + \sum_{s\geq k+1}\pi_s \tag{4.1.12}$$

for $k \geq 0$ and $\pi_{-1} \equiv 0$. We are assuming each GTP-T monomer in the cap is equally likely to hydrolyze. Thus, the last term on the right-hand side represents the probability that a cap of length greater than k hydrolyzes at the $k+1$th site. Adding the first $k-1$ equations gives (in steady state)

$$\pi_1 + 2\pi_2 + \ldots + (k-1)\pi_{k-1} + \lambda\pi_{k-1} = N_1 + N_2 + \ldots + N_k,$$

where $N_k = \sum_{s\geq k}\pi_s$. Using the fact that $N_j = N_k + \pi_{k-1} + \ldots + \pi_j$ for $j < k$, we see that

$$\pi_{k-1} = \frac{k}{\lambda}N_k.$$

From the identity $N_{k-1} - N_k = \pi_{k-1}$, it follows that

$$N_k = \frac{\lambda}{k+\lambda}N_{k-1}.$$

Iterating this equation and using $N_0 \equiv 1$ give

$$N_k = \frac{\lambda^k \Gamma(1+\lambda)}{\Gamma(k+1+\lambda)},$$

where $\Gamma(z)$ is the gamma function

$$\Gamma(z) = \int_0^\infty t^{z-1} \mathrm{e}^{-t} dt, \tag{4.1.13}$$

with

$$\frac{\Gamma(1+\lambda)}{\Gamma(k+1+\lambda)} = \frac{1}{(k+\lambda)(k-1+\lambda)\ldots(1+\lambda)}.$$

Thus, the stationary cap length distribution is

$$\pi_k = \frac{(k+1)\lambda^k \Gamma(1+\lambda)}{\Gamma(k+2+\lambda)}. \tag{4.1.14}$$

We can now calculate the mean cap length using

$$\langle k \rangle = \sum_{k\geq 1} k\pi_k = \sum_{k\geq 1} k[N_k - N_{k+1}] = \sum_{k\geq 1} N_k = -1 + \sum_{k\geq 0} N_k.$$

Given the solution for N_k and the properties of confluent hypergeometric functions $F(a,c;x)$ (see **Box 4A**), we have

$$\langle k \rangle = -1 + F(1, \lambda+1; \lambda) = -1 + \lambda \mathrm{e}^{\lambda} \lambda^{-\lambda} \gamma(\lambda, \lambda),$$

where $\gamma(a,x)$ is the incomplete gamma function. Finally, using the asymptotic result

$$\gamma(\lambda,\lambda) \to \sqrt{\frac{\pi}{2\lambda}} \lambda^{\lambda} \mathrm{e}^{-\lambda}, \quad \lambda \gg 1,$$

we see that

$$\langle k \rangle \to \sqrt{\pi\lambda/2}. \tag{4.1.15}$$

Thus a growing microtubule with λ GTP-T monomers has a cap size that scales as $\sqrt{\lambda}$.

Box 4A. Hypergeometric series.

The confluent hypergeometric function is defined according to the infinite series

$$F(a,c;x) = \sum_{k=0}^{\infty} \frac{(a)_n x^n}{(c)_n n!},$$

where we have used the Pochhammer symbol

$$(a)_n = a(a+1)(a+2)\ldots(a+n-1), \quad (a)_0 = 1.$$

The congruent hypergeometric function has the integral representation

$$F(a,c;x) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)} \int_0^1 \mathrm{e}^{xt} t^{a-1}(1-t)^{c-a-1} dt$$

for $\mathrm{Re}[c] > \mathrm{Re}[a] > 0$. An important special case is

$$\begin{aligned} F(1,1+\lambda;x) &= \int_0^1 \mathrm{e}^{xt}(1-t)^{\lambda-1} dt = \mathrm{e}^x \int_0^1 \mathrm{e}^{-ux} u^{\lambda-1} du \\ &= \mathrm{e}^x x^\lambda \int_0^x \mathrm{e}^{-u} u^{\lambda-1} du \\ &= \lambda \mathrm{e}^x x^{-\lambda} \gamma(\lambda, x), \end{aligned}$$

where $\gamma(\lambda,x)$ is the incomplete gamma function.

Constrained Growth. There are two mechanisms for slowing the growth rate; conversion from GTP-T to GTD-T at the tip resulting in a reduced rate of attachment (for $p < 1$) and detachment of GTD-T at the tip ($\mu > 0$). First, consider the effect of having $p < 1$ but no detachment. In determining the rate of growth, it is now necessary to keep track of the hydrolysis state of the end monomer. Thus, the kinetic equation for the number of GTP-T monomers in the chain becomes

$$\frac{dx}{dt} = -x + p\lambda\pi_0 + \lambda(1-\pi_0),$$

where π_0 is the probability that there is no cap. Extending Eq. (4.1.12) to the case $p < 1$ gives

$$\frac{d\pi_0}{dt} = -p\lambda\pi_0 + (1-\pi_0).$$

This pair of equations yields the steady-state solution

$$n_0 = p\lambda \frac{1+\lambda}{1+p\lambda}.$$

The steady-state speed of growth of the microtubule is

$$V(p,\lambda) = p\lambda n_0 + \lambda(1-n_0) = p\lambda \frac{1+\lambda}{1+p\lambda}.$$

Let us now calculate the probability distribution of microtubule lengths $P(L,t)$. Let $X(L,t)$ and $Y(L,t)$ denote the conditional probabilities that the length equals L and the end monomer is GTP-T and GTD-T, respectively. These probabilities evolve according to

$$\frac{dX(L,t)}{dt} = \lambda X(L-1,t) + p\lambda Y(L-1,t) - (1+\lambda)X(L),$$
$$\frac{dY(L,t)}{dt} = X(L) - p\lambda Y(L).$$

Adding this pair of equations and using $P = X + Y$,

$$\frac{dP(L,t)}{dt} = \lambda X(L-1,t) + p\lambda Y(L-1,t) - \lambda X(L) - p\lambda Y(L).$$

For sufficiently long filaments, the state of the last monomer does not depend on polymer length so that

$$X(L) = (1-\pi_0)P(L), \quad Y(L) = \pi_0 P(L).$$

Substituting into the equation for dP/dt gives

$$\frac{dP(L,t)}{dt} = V(p,\lambda)[P(L-1,t) - P(L,t)]. \tag{4.1.16}$$

This represents a Poisson process with rate V so that

$$P(L,t) = \frac{(Vt)^L}{L!}\mathrm{e}^{-Vt}, \tag{4.1.17}$$

and

$$\langle L\rangle = Vt, \quad \mathrm{var}[L] = Vt.$$

Finally, consider the general case $\mu > 0$ and $p < 1$. The probability that there is no cap (the end monomer is GTD-T) evolves as

$$\frac{d\pi_0}{dt} = -p\lambda\pi_0 + (1-\pi_0) - \mu\mathscr{N}_0,$$

where $\mathscr{N}_0 = \Pr\{+-\rangle\}$, and the corresponding speed of growth is

$$V(\lambda,\mu,p) = p\lambda\pi_0 + \lambda(1-\pi_0) - \mu\pi_0.$$

The difficulty in analyzing the general case is due to fact one has to solve an infinite hierarchy of higher-order correlations in order to determine $\mathscr{N}_0$. However, progress can be made in certain limiting cases [11]. For example, suppose that $\lambda, \mu \ll 1$ so that rate of conversion GTP-T $\to$ GTD-T is much faster than the other processes. Consequently, hydrolysis occurs as soon as a monomer attaches, that is, $\pi_0 \approx 1$. Hence, the growth phase occurs when $p\lambda > \mu$ and

$$V = p\lambda - \mu.$$

On the other hand, when $\lambda \gg 1$, there is a high probability that the end monomer is GTP-T so that $\pi_0 \approx \mathcal{N}_0$. Consequently, the steady-state distribution is

$$\pi_0 = \frac{1}{1+p\lambda+\mu},$$

and

$$V = \lambda - \frac{(1-p)\lambda+\mu}{1+p\lambda+\mu}.$$

4.2 Brownian Motion in a Periodic Potential

One qualitative method for modeling the stepping of molecular motors is based on the theory of Brownian ratchets (see the extensive review by Reimann [529]). Here, we develop the theory by considering the classical problem of how to solve the FPE for a Brownian particle in a periodic potential. Therefore, consider the 1D FP equation

$$\frac{\partial p}{\partial t} = D_0 \left[\frac{1}{k_B T} \frac{\partial [V'(x) - F_0] p}{\partial x} + \frac{\partial^2 p}{\partial x^2} \right], \tag{4.2.1}$$

where $V(x)$ is an L-periodic potential, $V(x+L) = V(x)$ for all x, and F_0 is a constant external force (see Fig. 4.8). We begin by describing the standard Stratonovich-based calculation of the mean velocity [253, 529, 612, 701] and show that it is zero when $F_0 = 0$, i.e., the motor cannot do any useful work. We then consider one mechanism for breaking periodicity that is based on rectification. An alternative mechanism, involving ATP hydrolysis and the breaking of detailed balance, will be the

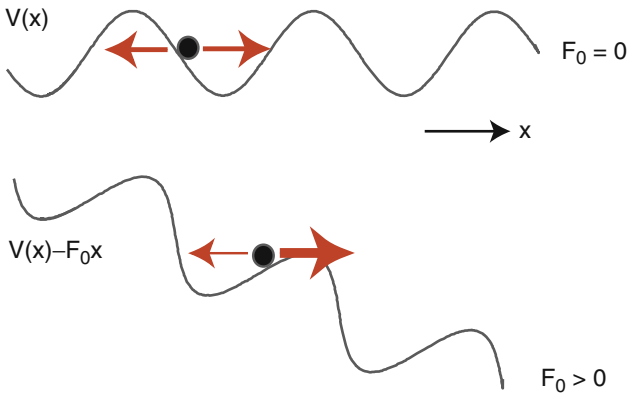


Fig. 4.8: Brownian particle moving in a periodic potential $V(x)$. In the absence of tilt ($F_0 = 0$) the mean velocity in the long-time limit is zero. On the other hand, in the presence of a tilt ($F_0 \neq 0$), the net motion of the particle is in the direction of the force

subject of Sect. 4.3. The first step is to introduce the effective potential or free energy $\mathscr{V}(x) = V(x) - F_0 x$ and to note that $\mathscr{V}'(x)$ is periodic even though $\mathscr{V}$ is not. Next we define the reduced probability density and currents

$$\hat{p}(x,t) = \sum_{n=-\infty}^{\infty} p(x+nL,t), \quad \hat{J}(x,t) = \sum_{n=-\infty}^{\infty} J(x+nL,t) \tag{4.2.2}$$

with

$$J(x,t) = -D_0 \left[\frac{1}{k_B T} \mathscr{V}'(x) p + \frac{\partial p}{\partial x} \right].$$

It immediately follows that

$$\hat{p}(x+L,t) = \hat{p}(x,t), \quad \int_0^L \hat{p}(x,t)dx = 1. \tag{4.2.3}$$

The periodicity of $\mathscr{V}'(x)$ implies that if $p(x,t)$ is a solution of the FP equation, then so is $p(x+nL,t)$. (Note that $p(x,t)$ itself is not periodic; otherwise it would not be possible to satisfy the normalization condition $\int_{\infty}^{\infty} p(x,t)dx = 1$.) The principle of superposition for a linear PDE then shows that $\hat{p}$ satisfies the FP equation

$$\frac{\partial \hat{p}(x,t)}{\partial t} + \frac{\partial \hat{J}(x,t)}{\partial x} = 0, \tag{4.2.4}$$

with

$$\hat{J}(x,t) = -D_0 \left[\frac{1}{k_B T} \mathscr{V}'(x) \hat{p} + \frac{\partial \hat{p}}{\partial x} \right] \tag{4.2.5}$$

and periodic boundary conditions at $x = 0, L$. There exists a stationary solution $\hat{p}_0$ of the reduced FP equation with constant flux $\hat{J}_0$ such that

$$\frac{d}{dx} \left(\mathrm{e}^{\mathscr{V}(x)/k_B T} \hat{p}_0(x) \right) = -\frac{\hat{J}_0}{D_0} \mathrm{e}^{\mathscr{V}(x)/k_B T}. \tag{4.2.6}$$

(The full FP equation does not have a nontrivial steady state, since $p(x,t) \to 0$ as $t \to \infty$.) Integrating this equation from x to $x+L$ and using periodicity yield the stationary solution

$$\hat{p}_0(x) = \frac{\hat{J}_0 \mathscr{N}(x)}{\left[1 - \mathrm{e}^{-F_0 L/k_B T}\right]}, \tag{4.2.7}$$

where

$$\mathscr{N}(x) = \frac{1}{D_0} \mathrm{e}^{-\mathscr{V}(x)/k_B T} \int_x^{x+L} \mathrm{e}^{\mathscr{V}(y)/k_B T} dy. \tag{4.2.8}$$

Finally, $\hat{J}_0$ is determined by imposing the normalization condition on $\hat{p}_0$.

A quantity of particular interest is the ensemble averaged velocity v with $vdt = \langle dX(t) \rangle$. It turns out that this is equal to the rate of change of the ensemble averaged position [529]. Recall that the solution of the FP equation (4.2.1), $p(x,t)$, is the probability density on the sample space Ω of solutions to the Langevin equation

$$dX(t) = \frac{F_0 - V'(x)}{\gamma} + \sqrt{2D_0}dW(t),$$

where γ is the drag coefficient with $D_0\gamma = k_BT$. (We assume some fixed initial condition $X(0) = x_0$). The connection between the two paradigms can be expressed as

$$p(x,t) = \langle \delta(x - X(t)) \rangle,$$

where for fixed (x,t), $\langle \ldots \rangle$ denotes averaging with respect to realizations of the Wiener process. (This should be contrasted with the definition $\langle x(t) \rangle = \int xp(x,t)dx$.) Taking differentials of both sides with respect to time gives

$$\partial_t p(x,t)dt = -\langle \delta'(x - X(t))dX(t) \rangle,$$

and, since $\partial_t p = -\partial_x J(x,t)$, implies that

$$J(x,t)dt = \langle \delta(x - X(t))dX(t) \rangle.$$

Integrating both sides with respect to x yields the result

$$\langle dX(t) \rangle = \left[\int_{-\infty}^{\infty} J(x,t)dx \right] dt. \tag{4.2.9}$$

The right-hand side of Eq. (4.2.9) can be rewritten as

$$-\left[\int_{-\infty}^{\infty} x\partial_x J(x,t)dx \right] dt = \left[\int_{-\infty}^{\infty} x\partial_t p(x,t)dx \right] dt = \frac{d\langle x(t) \rangle}{dt} dt.$$

We deduce the important result that [529]

$$\langle dX(t) \rangle = \frac{d\langle x(t) \rangle}{dt} dt.$$

Equation (4.2.9) thus implies that

$$v = \int_{-\infty}^{\infty} J(x,t)dx = \int_0^L \hat{J}(x,t)dx. \tag{4.2.10}$$

Since $v = L\hat{J}_0$ for constant current, it follows that

$$v = L \frac{1 - \mathrm{e}^{-F_0 L/k_B T}}{\int_0^L \mathscr{N}(x)dx}. \tag{4.2.11}$$

It can be seen that there is no net motion in a purely periodic potential, since the numerator vanishes when $F_0 = 0$. Moreover the net direction of motion for $F_0 \neq 0$ is in the direction of the applied force. Note that in the case of a space-dependent

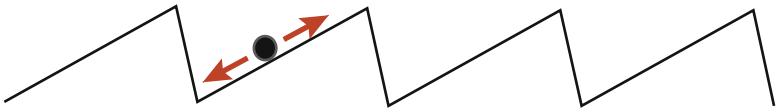


Fig. 4.9: Brownian particle moving in a periodic ratchet potential $V(x)$

diffusion coefficient $D(x)$, the above analysis is easily extended with $\mathscr{N}(x)$ now given by [82]

$$\mathscr{N}(x) = \mathrm{e}^{-\mathscr{V}(x)/k_BT} \int_x^{x+L} \frac{1}{D(y)} \mathrm{e}^{\mathscr{V}(y)/k_BT} dy.$$

The result that there is no net motion in a periodic potential ($F = 0$) can be counter-intuitive when considering ratchet potentials as shown in Fig. 4.9, since one might think that it is more difficult to move backward and cross the steep slope.

4.2.1 Polymerization Ratchet

One interesting application of ratchet potentials is to the so-called polymerization ratchet [507], which is a simplified model of the role of actin polymerization in changing the shape of a cell's membrane during cell motility [444, 447] (see Sect. 8.3). Suppose that a section of cell membrane wall is undergoing Brownian

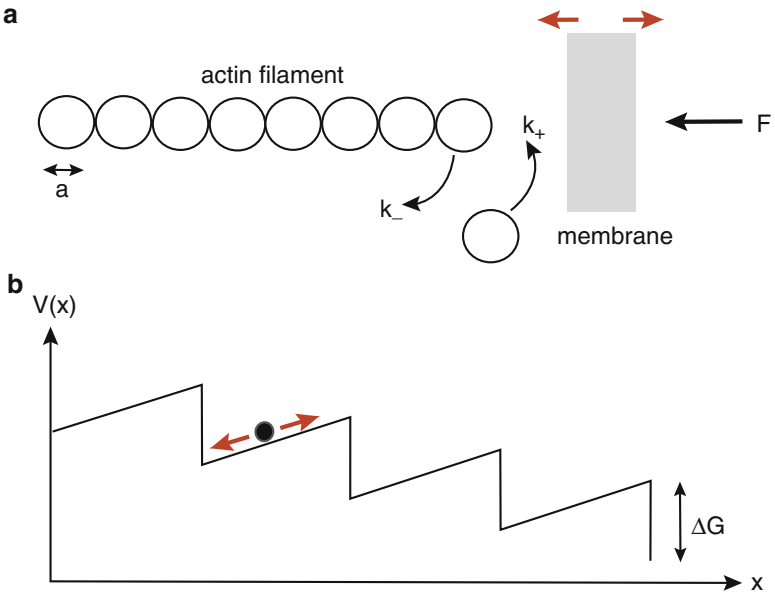


Fig. 4.10: (**a**) Cartoon of polymerization ratchet model. (**b**) Simplified ratchet model

motion in the presence of a resistive force F due to stretching (see Fig. 4.10a). This motion is rectified by the addition of actin monomers to the end of an actin polymer filament, whenever the gap x between membrane wall and filament is sufficiently large. Assume that in the absence of a load force, actin monomers are added at a rate $k_+ m$ and lost at a rate k_-, where m is the background concentration of monomers. First, consider the limiting case in which the mean time between attachments is sufficiently large so that the Brownian particle reaches thermal equilibrium. This means that the probability density for a gap of size x is given by the Boltzmann–Gibbs distribution (Sect. 1.4):

$$p(x) = \frac{F}{k_B T} \mathrm{e}^{-Fx/k_B T}.$$

An estimate of the mean polymerization velocity is then

$$v = a\left[k_+ m P(x > a) - k_-\right],$$

where a is the size of a monomer and

$$P(x > a) = \int_a^\infty p(x)dx = \mathrm{e}^{-Fa/k_B T},$$

Finally, using detailed balance,

$$\frac{k_+ m}{k_-} = \mathrm{e}^{\Delta G/k_B T},$$

where ΔG is the binding energy, we have

$$v = a k_- \left[\mathrm{e}^{[\Delta G - Fa]/k_B T} - 1\right], \tag{4.2.12}$$

which suggests that growth stops when the resistive force F becomes sufficiently large such that $F \geq F_S$, where the *stall force* $F_S = \Delta G/a$. A sketch of the velocity-load curve for typical values of k_+, m, k_-, and a is shown in Fig. 4.11.

Let us now turn to the diffusion-limited case, which has been analyzed by Peskin et al. [507] using a Fokker–Planck description of the process shown in Fig. 4.10a. Here, we will consider a reduced model, consisting of a Brownian particle moving in a ratchet potential (see Fig. 4.10b). This is obtained by ignoring spontaneous unbinding of monomers ($k_- = 0$) and assuming that as soon as the distance between the polymer and the wall is equal to a, a new monomer is immediately inserted, resulting in a sudden drop in energy by an amount ΔG. However, it is possible to reverse direction by jumping over a free energy barrier of height ΔG—this represents the dislodging of a monomer due to wall motion.

The analysis of the reduced model proceeds along similar lines to the general motion of a Brownian particle in a tilted potential with

$$\mathscr{V}(x) = Fx - n\Delta G, \quad na < x < (n+1)a.$$

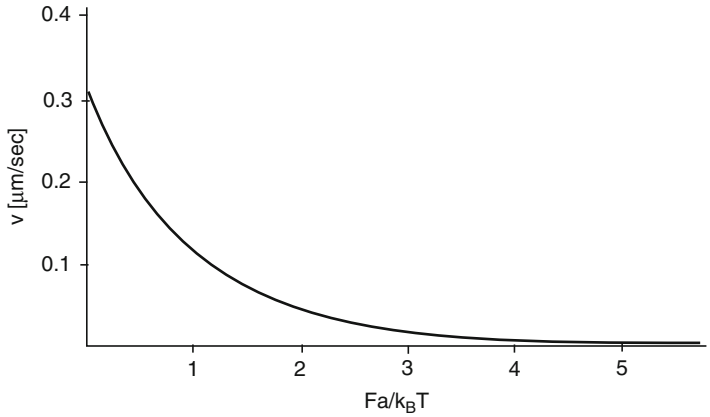


Fig. 4.11: Sketch of typical velocity–force curve based on Eq. (4.2.9)

However, one now needs to take into account the discontinuities in $V(x)$ at the points $x = na$, integer n. Thus, Eq. (4.2.6) still holds, but care must be taken when integrating this equation with respect to $x \in (0, a]$. That is, it is necessary to introduce the matching condition

$$\lim_{x \to a^+} \hat{p}_0(x) \mathrm{e}^{\mathscr{V}(x)} = \lim_{x \to a^-} \hat{p}_0(x) \mathrm{e}^{\mathscr{V}(x)}.$$

One finds that (see Ex. 4.5)

$$v = \frac{2D_0}{a} \frac{\omega^2/2}{\mathscr{A}(1 - \mathrm{e}^{-\omega}) - \omega}, \quad \omega = \frac{Fa}{k_B T} \tag{4.2.13}$$

with

$$\mathscr{A} = \frac{\mathrm{e}^{\Delta G/k_B T} - 1}{\mathrm{e}^{(\Delta G - Fa)/k_B T} - 1}. \tag{4.2.14}$$

Note that $v \to 0$ as $Fa \to \Delta G$, since $\mathscr{A} \to \infty$. On the other hand, in the regime $\Delta G \gg Fa$ and $k_B T \gg Fa$,

$$v \approx 2D_0/a.$$

This latter result can be understood as follows: in the absence of a force F, the mean time for a diffusive displacement of size a is $T = a^2/2D_0$ so that the mean speed is $v = a/T$.

4.2.2 Translocation Ratchet

Following gene expression, many proteins have to translocate into or across a cellular membrane. Examples include translocation through nuclear pores and through pores in the endoplasmic reticulum. It has been suggested that translocation may be

driven by a Brownian ratchet [212, 507, 592]. The basic mechanism is illustrated in Fig. 4.12. Once the protein chain enters a pore, thermal fluctuations cause it to diffuse back and forth through the pore without any net displacement. However, suppose that the protein has ratchet sites that are equally spaced along the chain with nearest neighbor separation δ. In the case of a perfect ratchet, it is assumed that once a ratchet site has passed through the pore it cannot reenter the pore, that is, it is reflected. On the other hand, for an imperfect ratchet there is a certain probability π of reflection. The latter could be due to the binding of a macromolecule (chaperonin) to the ratchet site on the distal side of the pore.

Consider a translocation ratchet and let $p(x,t)$ be the probability density that $X(t) = x$, where $X(t)$, $0 < X(t) < \delta$, is the position of the first ratchet site to the right of the pore exit. Let F be the net force resisting translocation of the protein. The corresponding FP equation takes the form

$$\frac{\partial p}{\partial t} + \frac{\partial J}{\partial x} = 0, \quad J = -\frac{DF}{k_B T} p - D\frac{\partial p}{\partial x}. \tag{4.2.15}$$

The corresponding boundary conditions for a perfect ratchet are

$$J(0,t) = J(\delta,t), \quad p(\delta,t) = 0. \tag{4.2.16}$$

The periodic flux condition expresses the fact that as soon as one ratchet site crosses $x = \delta$, another site appears at $x = 0$, with $x = \delta$ treated as an absorbing boundary. The steady-state solution satisfies the constant flux condition

$$-\frac{DF}{k_B T} p - D\frac{\partial p}{\partial x} = J_0.$$

Multiplying both sides by $D^{-1} e^{Fx/k_B T}$, integrating from x to δ, and using the absorbing boundary condition yields

$$p(x) = \frac{k_B T J_0}{DF} \left[e^{F(\delta - x)/k_B T} - 1 \right].$$

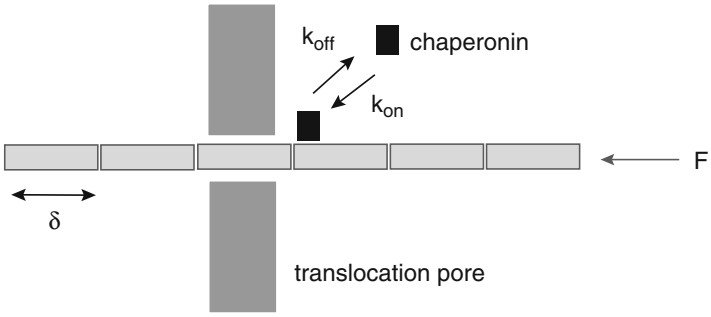


Fig. 4.12: Cartoon of a translocation ratchet

Imposing the normalization condition $\int_0^1 p(x)dx = 1$ then determines J_0 according to

$$1 = \frac{J_0\delta^2}{D}\frac{1}{\omega}\left[\mathrm{e}^{\omega} - 1 - \omega\right], \quad \omega = \frac{F\delta}{k_BT}.$$

It follows that the average speed of the perfect translocation ratchet is [507]

$$v = \delta J_0 = \frac{2D}{\delta}\frac{\omega^2}{\mathrm{e}^{\omega} - 1 - \omega}. \tag{4.2.17}$$

Now suppose that each ratchet site can exist in two states that are in equilibrium

$$S_0 \underset{k_{\rm off}}{\overset{k_{\rm on}}{\rightleftharpoons}} S_1,$$

with only S_1 ratcheted. Hence S_0 passes freely through the pore in both directions, whereas S_1 is reflected. The probability of being in the ratcheted state is then

$$\pi = \frac{k_{\rm on}}{k_{\rm on} + k_{\rm off}}.$$

The only modification of the perfect ratchet equations is that the absorbing boundary condition is replaced by [507]

$$p(\delta) = (1 - \pi)p(0). \tag{4.2.18}$$

Repeating the above calculation yields the modified velocity (see Ex. 4.6)

$$v = \delta J_0 = \frac{2D}{\delta}\left[\frac{\omega^2/2}{\dfrac{\mathrm{e}^{\omega} - 1}{1 - K(\mathrm{e}^{\omega} - 1)} - \omega}\right]. \tag{4.2.19}$$

Note that one major simplification of the above model is that it treats the translocating polymer as rigid. However, a polymer such as a protein or DNA tends to be highly coiled (small persistence length) so that one has to take into account an effective entropic force, reflecting the fact that a free polymer has many more configurational states than one that is threaded through a pore [117, 462, 493, 615]. The statistical mechanics of a polymer is considered in Sect. 4.5 and the application to translocation through a pore is developed in Sect. 7.3.4.

4.3 Brownian Ratchet Model of a Processive Molecular Motor

In performing a single step along a filament track, a molecular motor cycles through a sequence of conformational states before returning to its initial state (modulo the change in spatial location). Suppose that there is a total of M conformational states

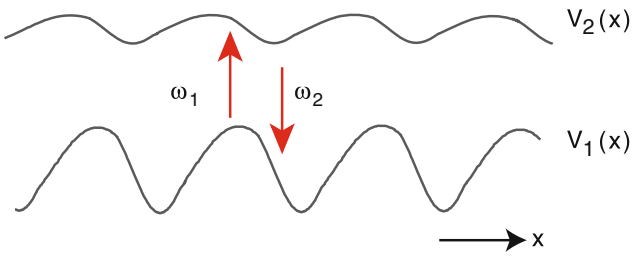


Fig. 4.13: Brownian ratchet model of a molecular motor that can exist in two internal states with associated l-periodic ratchet potentials $V_1(x)$ and $V_2(x)$. State transition rates are denoted by ω_1 and ω_2

in a single cycle labeled $i = 1,\ldots,M$. Given a particular state i, the motor is modeled as an overdamped, driven Brownian particle moving in an asymmetric periodic (ratchet) potential $V_i(x)$. A periodic potential is said to be *symmetric* if there exists Δx such that

$$V_i(-x) = V_i(x+\Delta x)$$

for all x, otherwise it is *asymmetric*. The asymmetry of the potentials reflects the fact that cytoskeletal filaments are polarized. The Langevin equation for the location of the particle $X(t)$, assuming that it remains in a given conformational state, is

$$dX = -\frac{V_i'(X)}{\gamma}dt + dW_i(t), \tag{4.3.20}$$

with $\langle dW_i(t)\rangle = 0$ and $\langle dW_i(t)dW_j(t')\rangle = 2D\delta_{i,j}\delta(t-t')dt\,dt'$. The corresponding FP equation is

$$\frac{\partial p_i(x,t)}{\partial t} = -\frac{\partial J_i(x,t)}{\partial x}, \tag{4.3.21}$$

where $p_i(x,t)$ is the probability density that the motor particle is in internal state i and at location x at time t and $J_i(x,t)$ is the probability flux

$$J_i(x,t) = \frac{1}{\gamma}\left[-V_i'(x) - k_BT\frac{\partial}{\partial x}\right]p_i(x,t), \tag{4.3.22}$$

where $D\gamma = k_BT$. If the state transitions between the conformational states are now introduced according to a discrete Markov process, then it is necessary to add source terms to the FP equation:

$$\frac{\partial p_i(x,t)}{\partial t} = -\frac{\partial J_i(x,t)}{\partial x} + \sum_{j\neq i}\left[\omega_{ij}(x)p_j(x,t) - \omega_{ji}(x)p_i(x,t)\right],$$

where $\omega_{ij}(x)$ is the rate at which the motor switches from state j to state i.

In order to develop the basic theory, consider the simple case of two internal states $N = 2$ following along the lines of [309, 495, 508, 516]. Then

$$\frac{\partial p_1(x,t)}{\partial t} + \frac{\partial J_1(x,t)}{\partial x} = -\omega_1(x)p_1(x,t) + \omega_2(x)p_2(x,t) \tag{4.3.23a}$$

$$\frac{\partial p_2(x,t)}{\partial t} + \frac{\partial J_2(x,t)}{\partial x} = \omega_1(x)p_1(x,t) - \omega_2(x)p_2(x,t). \tag{4.3.23b}$$

Note that adding the pair of equations together and setting $p = p_1 + p_2$, $J = J_1 + J_2$ leads to the conservation equations $\partial_t p + \partial_x J = 0$. An example of l-periodic ratchet potentials $V_1(x), V_2(x)$ is shown in Fig. 4.13, with l the basic step length of a cycle along the filament track. The analysis of the two-state model proceeds along similar lines to the one state model considered in Sect. 4.2. That is, set

$$\hat{p}_j(x,t) = \sum_{n=-\infty}^{\infty} p_j(x+nl,t), \quad \hat{J}_j(x,t) = \sum_{n=-\infty}^{\infty} J_j(x+nl,t). \tag{4.3.24}$$

The total probability flux can then be written as

$$\hat{J}(x,t) = -\frac{1}{\gamma}\left[V_1'(x)\hat{p}_1(x,t) + V_2'(x)\hat{p}_2(x,t) + k_BT\frac{\partial \hat{p}(x,t)}{\partial x}\right].$$

Consider the steady-state solution for which there is a constant total flux $\hat{J}_0$ so that

$$V_1'(x)\hat{p}_1(x) + V_2'(x)\hat{p}_2(x) + k_BT\frac{\partial \hat{p}(x)}{\partial x} = -\hat{J}_0\gamma.$$

Defining $\lambda(x) = \hat{p}_1(x)/\hat{p}(x)$, this equation can be rewritten as

$$V_{\rm eff}'(x)\hat{p}(x) + k_BT\frac{\partial \hat{p}(x)}{\partial x} = -\hat{J}_0\gamma, \tag{4.3.25}$$

where

$$V_{\rm eff}(x) = \int_0^x \left[\lambda(y)V_1'(y) + (1-\lambda(y))V_2'(y)\right] dy. \tag{4.3.26}$$

Suppose that the system is in thermodynamic equilibrium. The state transition rates and steady-state probabilities then satisfy the detailed balance condition (see Sect. 1.4)

$$\frac{\omega_1(x)}{\omega_2(x)} = {\rm e}^{[V_1(x)-V_2(x)]/k_BT} = \frac{\hat{p}_2(x)}{\hat{p}_1(x)}. \tag{4.3.27}$$

Therefore,

$$\lambda(x) = \frac{1}{1+{\rm e}^{-[V_1(x)-V_2(x)]/k_BT}}, \tag{4.3.28}$$

and, in particular, $\lambda(x)$ reduces to an l-periodic function. It follows that $V_{\rm eff}(x)$ in Eq. (4.3.25) is also an l-periodic potential and hence there is no net motion in a particular direction (in the absence of an external force or tilt) (see Sect. 4.2). In conclusion, in order for a molecular motor to sustain directed motion that can pull against

an applied load, we require a net positive supply of chemical energy that maintains the state transition rates away from detailed balance—this is the role played by ATP.

Therefore, consider the situation in which transitions between the two states occur as a result of chemical reactions involving ATP hydrolysis. Denoting the two conformational states of the motor by M_1, M_2, the scheme is taken to be [495]

$$\begin{aligned}
&\mathrm{ATP} + M_1 \underset{\alpha_2}{\overset{\alpha_1}{\rightleftharpoons}} M_2 + \mathrm{ADP} + P \\
&\mathrm{ADP} + P + M_1 \underset{\gamma_2}{\overset{\gamma_1}{\rightleftharpoons}} M_2 + \mathrm{ATP} \\
&M_1 \underset{\beta_2}{\overset{\beta_1}{\rightleftharpoons}} M_2
\end{aligned}$$

with $\alpha_j, \gamma_j, \beta_j$ x-dependent. The first reaction pathway involves ATP hydrolysis with chemical free energy gain $\Delta\mu$ and a corresponding transition from state 1 to state 2, the second involves hydrolysis in the opposite direction, while the third involves thermal state transitions without any change in chemical free energy. Basic chemical kinetics implies that

$$\begin{aligned}
&\frac{\alpha_1}{\alpha_2} = \mathrm{e}^{(V_1 - V_2 + \Delta\mu)/k_B T}, \quad \frac{\gamma_1}{\gamma_2} = \mathrm{e}^{(V_1 - V_2 - \Delta\mu)/k_B T}, \\
&\frac{\beta_1}{\beta_2} = \mathrm{e}^{(V_1 - V_2)/k_B T}.
\end{aligned} \tag{4.3.29}$$

It follows that the net transition rates between the two conformational states are

$$\omega_1 = \alpha_2 \mathrm{e}^{(V_1 - V_2 + \Delta\mu)/k_B T} + \gamma_2 \mathrm{e}^{(V_1 - V_2 - \Delta\mu)/k_B T} + \beta_2 \mathrm{e}^{(V_1 - V_2)/k_B T} \tag{4.3.30}$$

$$\omega_2 = \alpha_2 + \gamma_2 + \beta_2. \tag{4.3.31}$$

Clearly detailed balance no longer holds. In general, it is now necessary to determine the steady-state solution of the pair of Eq. (4.3.23) numerically. Given such a solution, the efficiency of the motor doing work against a load F may be determined as follows. First the flux (4.3.22) has an additional term of the form $F p_i(x,t)/\gamma$. The mechanical work done per unit time against the external force is then $\dot{W} = Fv$ where $v = l\hat{J}_0$ is the velocity of the motor. On the other hand, the chemical energy consumed per unit time is $\dot{Q} = r\Delta\mu$, where r is the steady-state rate of ATP consumption:

$$r = \int_0^l [(\alpha_1(x) - \gamma_1(x))\hat{p}_1(x) - (\alpha_2(x) - \gamma_2(x))\hat{p}_2(x)]dx.$$

The efficiency of the motor is then defined to be [309] $\eta = Fv/r\Delta\mu$.

A major mathematical challenge is to determine how the effective speed of the molecular motor depends on the asymmetries of the potentials and/or the transition rates. (If these functions are symmetric, then there is no net polarization and the speed is zero.) In the weak diffusion limit, analytical tools from PDE theory and

transport processes have been used to prove the existence of a steady-state solution of Eq. (4.3.23) in a bounded domain with no-flux boundary conditions. Moreover, for certain classes of potential function and transition rates, the steady-state density localizes to one end or other of the domain—the so-called motor-effect [108, 506, 662]. A simple heuristic argument for directed motion can be given [309, 516] by considering the switch between an asymmetric ratchet potential $V_1(x)$ and a uniform potential $V_2(x) = \text{const}$ for which pure diffusion occurs (see Fig. 4.14). Suppose that the motor starts at a minimum of the potential $V_1(x)$ and is excited to state 2. In this state it undergoes diffusion, which generates a Gaussian probability density with a width $\sqrt{2Dt}$ at time t. The motor should spend sufficient time τ_2 in state 2 so that it has a reasonable chance to jump down to the well of the next minimum on the right and yet not enough time to jump too far to the left. This suggests that $\tau_2 \sim a^2/D$ where a is the width of the steep part of the potential. The motor also needs just enough time in state 1 in order to move down the shallow part of the potential to the next minimum. If the width of the shallow part is b and the maximum potential is V_1, then the net drift is $V_1/\gamma b$. Assuming that the drift induced by the force dominates diffusion in state 1, we have $\tau_1 \sim b^2\gamma/V_1 \ll \tau_2$. Such a condition violates detailed balance. One way to measure the deviation from detailed balance is to introduce the quantity

$$\Omega(x) = \omega_1(x) - \omega_2(x)\mathrm{e}^{[V_1(x)-V_2(x)]/k_BT}. \tag{4.3.32}$$

One finds that the mean velocity v depends on the amplitude $\Omega(x)$ and whether it is homogeneous (x-independent) or localized as shown in Fig. 4.14. In the homogeneous case, the motor speed is a unimodal function of the amplitude

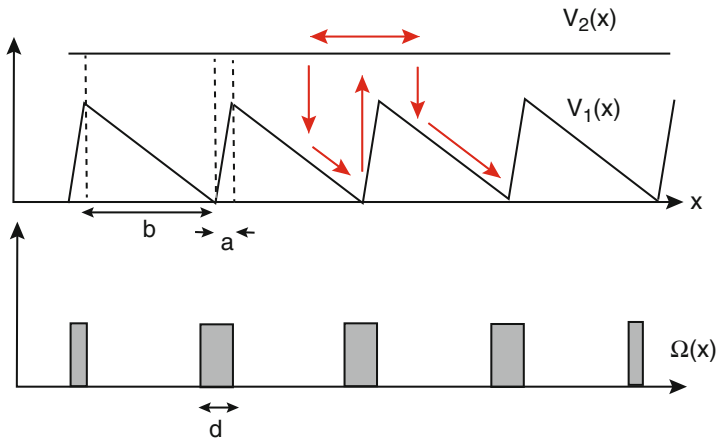


Fig. 4.14: Simplified Brownian ratchet model in which $V_1(x)$ is a periodic ratchet potential and $V_2(x)$ is constant. Also shown is an example of a localized function $\Omega(x)$ that signals regions where there is a breakdown of detailed balance

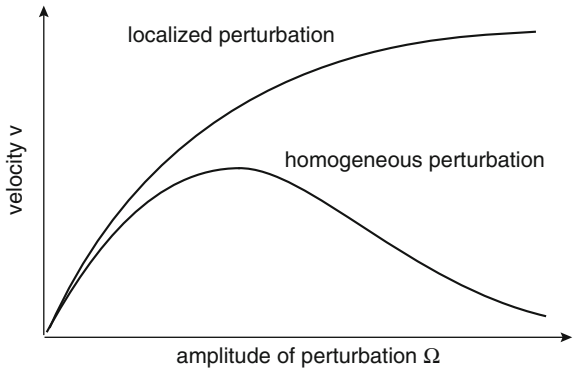


Fig. 4.15: Schematic diagram of the mean velocity v (for zero external force) of a two-state Brownian ratchet as a function of Ω, which measures the departure from equilibrium and is related to the fuel concentration. (Redrawn from [309].)

$\Omega_0 = \max_x |\Omega(x)|$, with the maximum at $\Omega_0 \sim 1/\tau_1$ and $v \to 0$ as $\Omega_0 \to \infty$. On the other hand the speed is monotonically increasing for a localized perturbation (see Fig. 4.15).

The Brownian ratchet is one important example of a stochastic system with a *nonequilibrium steady state* (NESS). An NESS has a number of characteristics that distinguish it from an equilibrium state: irreversibility, breakdown of detailed balance, and free energy dissipation. In particular, it is a steady state in which there are constant nonzero fluxes or currents. For a recent review of the theory and applications of NESSs see Zhang et al. [255, 701]. Although Brownian ratchet models provide important insights into the mechanisms underlying molecular motor dynamics, they have certain limitations. First, as we have already highlighted, it is difficult to obtain analytical solutions of the full equations in order to construct velocity–force curves, for example. Second, there is currently not enough experimental data regarding the potentials $V_j(x)$ and transition rates $\omega_{ij}(x)$ to sufficiently constrain models. Moreover, a large number of model parameters are needed to specify these functions, making data fitting problematic.

The above motivates an alternative approach to modeling molecular motors, based on a discrete Markov process [165, 348, 390, 391]. The basic idea is to take the transition rate functions to be localized at a discrete set of spatial positions $x = x_k$, $k = 1, \ldots, K$, and to replace the continuum diffusion and drift terms by hopping rates between nearest lattice sites. The resulting discrete Brownian ratchet model can be mapped on to a stochastic network of KM states as shown in Fig. 4.16. The stochastic dynamics is now described by a master equation, an example of which is

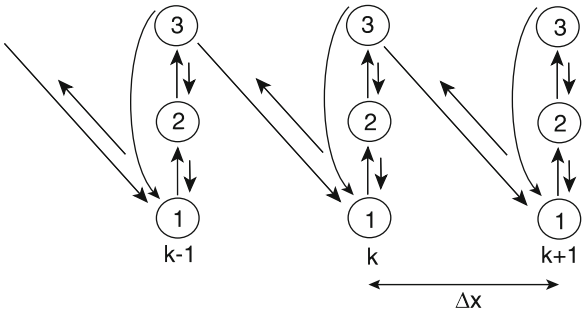


Fig. 4.16: State transition diagram for a discrete Brownian ratchet that cycles through $M = 3$ internal states and makes a single step of length Δx

$$\frac{dP_{km}(t)}{dt} = \sum_{n \neq m} [P_{kn}(t)W_{km;kn} - P_{km}(t)W_{kn;km}] + P_{k+1,1}(t)W_{kM;k+1,1} \qquad (4.3.33)$$
$$+ P_{k-1,M}(t)W_{k1;k-1,M} - P_{k,1}(t)W_{k-1,M;k,1} - P_{k,M}(t)W_{k+1,1;k,M},$$

where $P_{km}(t) = p_m(x_k, t)$ and for "vertical" transitions $W_{km;kn} = \omega_{mn}(x_k)$. In this example steps along the filament (power strokes) only occur between states $m = 1$ and $m = M$. One can then use methods developed by Derrida [137] to calculate the effective diffusion and velocity of a particle whose state probability evolves

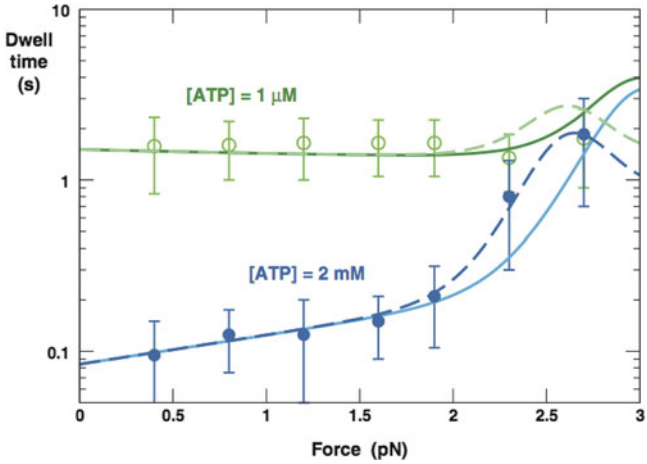


Fig. 4.17: Mean dwell times of myosin V as a function of external load at different ATP concentrations. The symbols correspond to the experimental data of Mehta et al. [433], whereas the solid lines are theoretical predictions from a discrete stochastic model analyzed by Kolomeisky and Fisher [347]. (Figure adapted from Kolomeisky and Fisher [347].)

according to such a master equation. One of the advantages of the discrete models is that the relatively small number of parameters makes it easier to fit to experimental data such as shown in Fig. 4.3 [348]. This is illustrated in Fig. 4.17, which shows the results of fitting a discrete model to data on the ATP dependence of myosin V dwell times, that is, the times between successive steps.

4.4 Collective Effects of Multiple Molecular Motors

In many cases, molecular motors work in groups rather than in isolation [245]. For example, the number of myosin motors involved in muscle contraction can be around 10^{19}, the number of dynein motors responsible for the beating of cilia or flagella is roughly 10^4, and up to ten coordinated motors can be involved in the intracellular transport of vesicles. A useful characterization of the collective behavior in motor assemblies distinguishes between *rowers* and *porters* [379]. Rowers spend most of their time unbound from their cytoskeletal filaments so they need to operate as part of a large assembly in order to produce sufficiently high velocities. This is exemplified by various classes of myosin motor. On the other hand, porters such as kinesin need to be more carefully coordinated, since the presence of other motors can impede the motion of any individual motor within the assembly. A variety of theoretical models have shown that there are interesting collective effects in motor ensembles, including bidirectional motion, spontaneous oscillations, hysteresis, and the formation of self-organizing structures (see also Chap. 8).

4.4.1 Intracellular Cargo Transport by Multiple Motors

Often intracellular cargo such as a vesicle is transported by multiple motors forming a motor/cargo complex. Here we will consider a simple model of a motor/cargo complex, in which there is only a single type of motor, kinesin, say responsible for transport [340]. Suppose that there are N identical motors irreversibly attached to the cargo particle, but which can bind to and unbind from the filament along which they move. Thus, the number n of motor molecules that are bound to the filament can vary between $n = 0$ and $n = N$. Hence, there are $N + 1$ different states of the cargo particle corresponding to the unbound state with $n = 0$ and to N bound states with $n = 1, 2, \ldots, N$. Each of these bound states contains $N!/(N-n)!n!$ substates corresponding to the different combinations of connecting n motor molecules to the filament. The dynamics of the motor complex depends on properties of individual motors combined with the observation that an applied force is shared equally between the motors.

(a) When bound to a microtubule, the velocity of a single molecular motor decreases approximately linearly with the force applied against the movement of the motor [658]. Thus, each motor is assumed to satisfy the linear force–velocity relation

$$v(F) = \begin{cases} v_f(1 - F/F_s) & \text{for } F \le F_s \\ v_b(1 - F/F_s) & \text{for } F \ge F_s, \end{cases} \tag{4.4.1}$$

where F is the applied force, F_s is the stall force satisfying $v(F_s) = 0$, v_f is the forward motor velocity in the absence of an applied force in the preferred direction of the particular motor, and v_b is the backward motor velocity when the applied force exceeds the stall force.

(b) The binding rate is independent of the applied force, whereas the unbinding rate is taken to be an exponential function of the applied force:

$$\pi(F) = \bar{\pi}, \quad \gamma(F) = \bar{\gamma} e^{\frac{F}{F_d}}, \tag{4.4.2}$$

where F_d is the experimentally measured force scale on which unbinding occurs. The force dependence of the unbinding rate is based on measurements of the walking distance of a single motor as a function of load [571], in agreement with Kramers' rate theory [253] (see Sect. 3.3).

(c) Now suppose that the externally applied load or force F acts on a motor/cargo complex with n independent molecular motors. If the motors are not directly coupled to each other, then they act independently and share the load. It follows that a single motor feels the force F/n. Hence, the velocity of the cargo when there are n bound motors is

$$v_n = v(F/n).$$

Moreover, Eq. (4.4.2) implies that the population binding and unbinding rates take the form

$$\gamma_n = n\gamma(F/n), \quad \pi_n = (N - n)\bar{\pi}. \tag{4.4.3}$$

One can model transitions between the different internal states of the cargo complex using a birth–death master equation (see Fig. 4.18). Let $P_n(t)$ be the probability that there are n bound motors at time t. Then

$$\frac{dP_n}{dt} = \gamma_{n+1}P_{n+1} + \pi_{n-1}P_{n-1} - (\gamma_n + \pi_n)P_n. \tag{4.4.4}$$

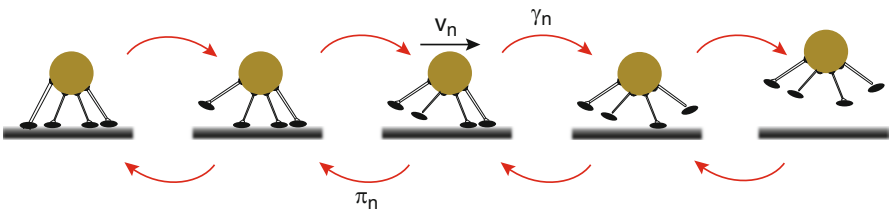


Fig. 4.18: Model of cooperative motor transport

There are a number of quantities that characterize properties of the transport process. We begin by calculating the steady-state distribution of bound motors, which satisfies the equation $J_n = J_{n+1}$ with

$$J_n = \gamma_n P_n - \pi_{n-1} P_{n-1}.$$

Since n is nonnegative, we have $P_n = 0$ for all $n < 0$, which means that $J_n = 0$ for all $n \geq 0$. Hence,

$$\gamma_{n+1} P_{n+1} = \pi_n P_n,$$

so that by iteration,

$$P_n = P_0 \prod_{i=0}^{n-1} \frac{\pi_i}{\gamma_{i+1}}, \tag{4.4.5}$$

with P_0 determined from the normalization $\sum_{n=0}^{N} P_n = 1$. The correctly normalized probability distribution of the bound states is then

$$\tilde{P}_n = \frac{P_n}{1 - P_0}.$$

It follows that the mean number of bound motors in steady state is

$$N_b = \sum_{n=1}^{N} \frac{n P_n}{1 - P_0}, \tag{4.4.6}$$

and the mean cargo velocity is

$$v_{\text{eff}} = \sum_{n=1}^{N} v_n \frac{P_n}{1 - P_0}. \tag{4.4.7}$$

The steady-state distribution of the number of bound motors also yields an explicit expression for the effective cargo detachment rate. In steady state, the effective cargo attachment and detachment rates satisfy

$$\gamma_{\text{eff}}(1 - P_0) = \pi_{\text{eff}} P_0,$$

where $1 - P_0$ is the probability that the cargo is bound to the filament via at least one motor. The effective binding rate is $\pi_{\text{eff}} = \bar{\pi}$, since attachment is established as soon as one motor binds to the filament. Thus

$$\gamma_{\text{eff}} = \frac{\bar{\pi} P_0}{1 - P_0} = \frac{\gamma_1 P_1}{1 - P_0}.$$

Using

$$P_{n+1} = P_1 \prod_{i=1}^{n} \frac{\pi_i}{\gamma_{i+1}}, \quad n \geq 1$$

and $1-P_0 = P_1 + \sum_{n=1}^{N-1} P_{n+1}$, we have

$$\gamma_{\rm eff} = \gamma_1 \left[1 + \sum_{n=1}^{N-1} \prod_{i=1}^{n} \frac{\pi_i}{\gamma_{i+1}} \right]^{-1}. \tag{4.4.8}$$

Explicit formulae can be obtained in the absence of a load force, for which $\gamma_n = n\bar{\gamma}$ and $\pi_n = (N-n)\bar{\pi}$ (see Ex. 4.7). For example, the steady-state distribution is

$$P_n = P_0 \frac{N!}{(N-n)!n!} \left(\frac{\bar{\pi}}{\bar{\gamma}} \right)^n, \quad P_0 = \left(1 + \frac{\bar{\pi}}{\bar{\gamma}} \right)^{-N}$$

and

$$N_b \approx N \frac{\bar{\pi}/\bar{\gamma}}{1 + \bar{\pi}/\bar{\gamma}}$$

for large N.

4.4.2 Tug-of-War Model

Using SPT (Sect. 1.2), trajectories of individual motor–cargo complexes can be recorded and shown to exhibit many random turning events [244, 359, 577, 677]. This immediately raises the issue of how bidirectional transport is achieved, given that motors such as kinesin and dynein are unidirectional. Recall that microtubules are polarized filaments with biophysically distinct + and − ends and the polarity determines the preferred direction of motion of individual motors: kinesin (dynein) moves towards the + (−) end. There is considerable debate in the literature regarding the most likely mechanism for bidirectional transport. Several different scenarios have been proposed including those shown in Fig. 4.19: (a) an asymmetric *tug-of-war model* involving the joint action of multiple kinesin and dynein motors pulling in opposite directions; (b) a symmetric ToW model where all the motors are of the same type, but they are distributed on microtubules of opposite polarity; (c) a hopping model, in which the whole motor–cargo complex hops between microtubules of opposite polarity. Yet another suggested mechanism (not shown) is some form of coordination complex that controls the switching between different motor species. It might be possible to apply statistical methods to SPT data which, when combined with knowledge of the individual motor dynamics, could identify the underlying mechanism(s) for bidirectional transport [14]. However, as far as we are aware, the debate continues! For the sake of illustration, we will focus on the first two models here.

It has been hypothesized that the experimentally observed bidirectional motion of molecular motors could be due to the joint action of multiple kinesin and dynein motors pulling in opposite directions. Suppose that a certain vesicular cargo is transported along a one-dimensional track via N_+ right-moving (anterograde) motors and N_- left-moving (retrograde) motors (see Fig. 4.19). At a given time t, the internal

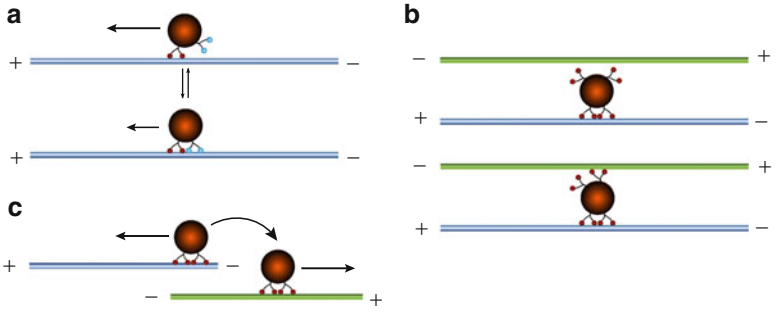


Fig. 4.19: Molecular motor-assisted models of bidirectional transport. (**a**) Asymmetric tug-of-war model: opposing groups of motors (e.g., dynein and kinesin) compete to transport along a single polarized microtubule track. (**b**) Symmetric tug-of-war model: groups of motors of the same directional preference are distributed among two parallel microtubules of opposite polarity. (**c**) Hopping model in which a motor–cargo complex jumps between microtubules of opposite polarity

state of the cargo–motor complex is fully characterized by the numbers n_+ and n_- of anterograde and retrograde motors that are bound to a microtubule and thus actively pulling on the cargo. Assume that over the time scales of interest all motors are permanently bound to the cargo so that $0 \leq n_\pm \leq N_\pm$. The ToW model of Muller et al. [456, 457] assumes that the motors act independently other than exerting a load on motors with the opposite directional preference. (However, some experimental work suggests that this is an oversimplification, i.e., there is some direct coupling between motors [150].) Thus the properties of the motor complex can be determined from the corresponding properties of the individual motors together with a specification of the effective load on each motor.

The ToW model can be constructed as a generalization of the cooperative motor model. At the single-motor level, Eqs. (4.4.1) and (4.4.2) still hold, except that the value of parameters such as the stall force will differ for kinesin and dynein. However, even in the absence of an externally applied load, there is now an effective force on one class of motor due to the opposing action of the other class. Let F_c denote the net load on the set of anterograde motors, which is taken to be positive when pointing in the retrograde direction. If the molecular motors are not directly coupled to each other, then a single anterograde motor feels the force F_c/n_-, whereas a single retrograde motor feels the opposing force $-F_c/n_+$. At the population level, the binding and unbinding rates for the two types of motor are

$$\gamma_j(n_j) = n_j \bar{\gamma}_j e^{F_c/F_{D,j} n_j}, \quad \pi_j(n_j) = (N_j - n_j)\bar{\pi}_j \quad j = \pm. \tag{4.4.9}$$

The cargo force F_c is determined self-consistently by the condition that all the motors move with the same cargo velocity v_c. Suppose that $N_+ \geq N_-$ so that the net motion is in the anterograde direction, which is taken to be positive. In this case,

the forward motors are stronger than the backward motors so that $n_+F_{s+} > n_-F_{s-}$. Equation (4.4.1) implies that

$$v_c = v_{f+}(1 - F_c/(n_+F_{s+})) = -v_{b-}(1 - F_c/(n_-F_{s-})). \tag{4.4.10}$$

This generates a unique solution for the load F_c and cargo velocity v_c:

$$F_c(n_+,n_-) = (\mathscr{F} n_+F_{s+} + (1-\mathscr{F})n_-F_{s-}), \tag{4.4.11}$$

where

$$\mathscr{F} = \frac{n_-F_{s-}v_{f-}}{n_-F_{s-}v_{f-} + n_+F_{s+}v_{b-}}, \tag{4.4.12}$$

and

$$v_c(n_+,n_-) = \frac{n_+F_{s+} - n_-F_{s-}}{n_+F_{s+}/v_{f+} + n_-F_{s-}/v_{b-}}. \tag{4.4.13}$$

The corresponding expressions when the backward motors are stronger, $n_+F_{s+} < n_-F_{s-}$, are found by interchanging v_f and v_b.

The original study of [456, 457] considered the stochastic dynamics associated with transitions between different internal states (n_+,n_-) of the motor complex, without specifying the spatial position of the complex along a 1D track. This defines a Markov process with a corresponding master equation for the time evolution of the probability distribution $P(n_+,n_-,t)$. They determined the steady-state probability distribution of internal states and found that the motor complex exhibited at least three different modes of behavior (see Fig. 4.20), which were consistent with experimental studies of motor transport using SPT (Sect. 1.2); the transitions between these modes of behavior depend on motor strength, which primarily depends upon the stall force.

(i) The motor complex spends most of its time in states with approximately zero velocity.
(ii) The motor complex alternates between fast backward and forward movements, so that there is a bimodal velocity distribution with peaks close to the single-motor velocities of $1\,\mu\mathrm{m\,s}^{-1}$.
(iii) The motor complex exhibits fast backward and forward movement interrupted by stationary pauses, which is consistent with experimental studies of bidirectional transport. The velocity distribution now has three peaks.

One of the useful features of the ToW model is that it allows various biophysical signaling mechanisms to be incorporated into the model [472, 473, 514]. This will be exploited in Sect. 7.6.4, when we use the ToW model to study the effects of local chemical signaling on intracellular cargo transport.

ATP Signaling. Experimentally, it is found that [ATP] primarily affects the stall force, forward motor velocity, and unbinding rate (see for example Fig. 4.3). There are a number of models of the [ATP] and force-dependent motor parameters that closely match experiments for both kinesin [183, 449, 571, 658] and

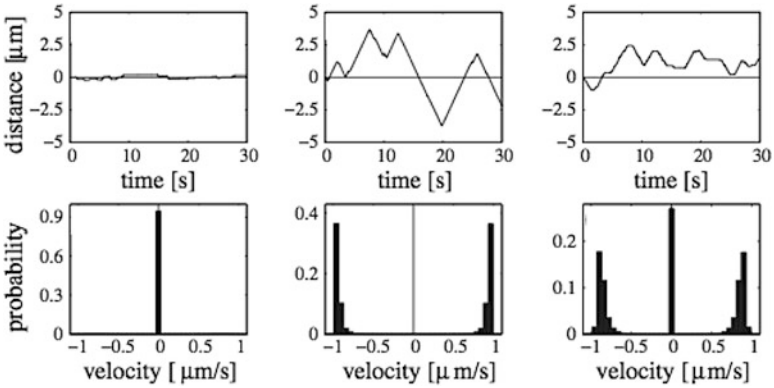


Fig. 4.20: Motility states for the symmetric tug-of-war model consisting of $N = 4$ plus and $N = 4$ minus motors. The three columns correspond to the three listed motility states (i), (ii), and (iii), respectively. The *top row* shows a typical trajectory for given motility state, whereas the *bottom row* shows a histogram of the distribution of velocities. The different motility behavior is obtained by taking different stall forces F_S and unbinding rates $\bar{\varepsilon}$: (i) $F_S = 2\,\mathrm{pN}, \bar{\varepsilon} = 0.4\,\mathrm{s}^{-1}$, (ii) $F_S = 6\,\mathrm{pN}, \varepsilon_0 = 1\,\mathrm{s}^{-1}$, (iii) $F_S = 4.75\,\mathrm{pN}, \varepsilon_0 = 0.4\,\mathrm{s}^{-1}$. These and other single-motor parameter values are based on experimental data from kinesin 1: detachment force $F_d = 3\,\mathrm{pN}$, binding rate $\bar{\pi} = 5\,\mathrm{s}^{-1}$, forward velocity $v_f = 1\,\mu\mathrm{m\,s}^{-1}$, and backward velocity $v_b = 6\,\mathrm{nm\,s}^{-1}$ (Adapted from Muller et al. [457])

dynein [201, 335]. We give some examples of [ATP]-dependent parameters. First, the forward velocity can be modeled using Michaelis–Menten kinetics:

$$v_f([\mathrm{ATP}]) = \frac{v_f^{max}[\mathrm{ATP}]}{[\mathrm{ATP}] + K_v}, \tag{4.4.14}$$

where $v_f^{max} = 1\,\mu\mathrm{m\,s}^{-1}$, $K_v = 79.23\,\mu\mathrm{M}$ for kinesin and $v_f^{max} = 0.7\,\mu\mathrm{m\,s}^{-1}$, $K_v = 38\,\mu\mathrm{M}$ for dynein. (The backward velocity of both kinesin and dynein is small, $v_b \approx \pm 0.006\,\mu\mathrm{m\,s}^{-1}$, so that the [ATP] dependence can be neglected.) The binding rate is determined by the time necessary for an unbound motor to diffuse within range of the microtubule and bind to it, which is assumed to be independent of [ATP]. The unbinding rate of a single motor under zero load can be determined using the [ATP]-dependent average run length $L_k([\mathrm{ATP}]) = L_k^{max}/([\mathrm{ATP}] + K_u)$. The mean time to detach from the microtubule is $v_f([\mathrm{ATP}])/L_k([\mathrm{ATP}])$ so that

$$\bar{\gamma}([\mathrm{ATP}]) = \frac{v_f^{max}([\mathrm{ATP}] + K_u)}{L_k^{max}([\mathrm{ATP}] + K_v)}, \tag{4.4.15}$$

where $L_k^{max} = 0.86\,\mu\mathrm{m}$, $K_u = 3.13\,\mu\mathrm{M}$ for kinesin and $L_k^{max} = 1.5\,\mu\mathrm{m}$, $K_u = 1.5\,\mu\mathrm{M}$ for dynein. Finally, a model for the [ATP]-dependent stall force of kinesin is

$$F_s([\mathrm{ATP}]) = F_s^0 + \frac{(F_s^{max} - F_s^0)[\mathrm{ATP}]}{K_s + [\mathrm{ATP}]}, \tag{4.4.16}$$

where $F_s^0 = 5.5\,\text{pN}$, $F_s^{max} = 8\,\text{pN}$, $K_s = 100\,\mu\text{M}$ for kinesin and $F_s^0 = 0.22\,\text{pN}$, $F_s^{max} = 1.24\,\text{pN}$, $K_s = 480\,\mu\text{M}$ for dynein.

Tau Signaling. The second signaling mechanism involves microtubule-associated proteins (MAPs). These molecules bind to microtubules and effectively modify the free energy landscape of motor–microtubule interactions [622]. For example, tau is a MAP found in the axon of neurons and is known to be a key player in Alzheimer's disease [351]. Experiments have shown that the presence of tau on the microtubule can significantly alter the dynamics of kinesin, specifically by reducing the rate at which kinesin binds to the microtubule [655]. It has also been shown that, at the low tau concentrations affecting kinesin, dynein is relatively unaffected by tau. Thus tau signaling can be incorporated into the ToW model by considering a tau concentration-dependent kinesin binding rate of the form [474]

$$\bar{\pi}(\tau) = \frac{\bar{\pi}^{max}}{1 + e^{-\gamma(\tau_0 - \tau)}}, \tag{4.4.17}$$

where τ is the dimensionless ratio of tau per microtubule dimer and $\bar{\pi}^{max} = 5\,\text{s}^{-1}$. The remaining parameters are found by fitting the above function to experimental data [655], so that $\tau_0 = 0.19$ and $\gamma = 100$.

4.4.3 Collective Extraction of Membrane Nanotubes

Another example of collective motor activity occurs in the extraction of a *membrane nanotube* or *tether* from a vesicle. Membrane nanotubes play an important role in lipid and protein exchange between various organelles of the early secretory pathways such as the ER and Golgi apparatus (see also Sect. 8.4) [609, 636]. They have also been observed in vitro, where the dynamical clustering of several molecular motors at the tip of a nanotube is required to pull it from the host vesicle [92, 352, 377]. The basic experimental setup involves a unilamellar vesicle coated with kinesin proteins, which are permanently attached to the membrane via their tail domains, and can bind/unbind to microtubules via their motor domains. The initial membrane tension τ_0 of the vesicle is determined by fixing the osmotic pressure difference between the interior and exterior of the vesicle. The kinesin-coated vesicle is sedimented on a network of microtubules fixed on a glass surface. Under suitable conditions, a membrane nanotube is formed when the kinesin motors bind microtubules and walk towards the plus end, deforming the vesicle membrane. Two types of behaviors are observed following formation of a nanotube. In the majority of cases, the membrane tubes simply stall at a certain length, whereas in the remaining cases the length of a tubed oscillates between two values; the oscillations are characterized by a slow growth phase and a fast retraction phase.

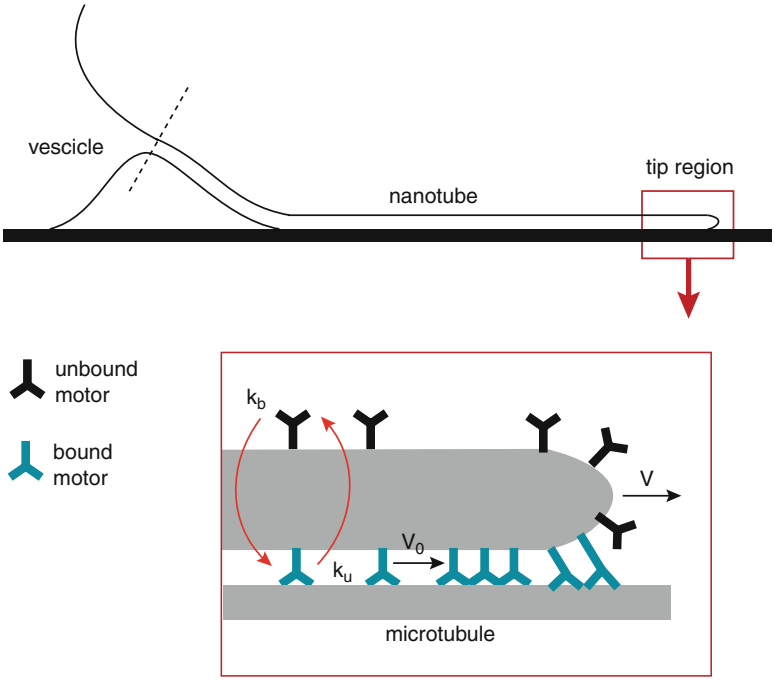


Fig. 4.21: Sketch of the extraction of a membrane nanotube from a vesicle by the collective action of kinesin molecular motors. A *side view* shows the different regions (vesicle, tube, and tip). The *inset* represents a scaled-up version of the tip region, where V is the velocity of the tube and of bound motors at the tip, V_0 is the zero-load velocity of bound kinesins, and k_u, k_b are the unbinding rate at zero load and the binding rate of kinesins onto MTs, respectively

The basic mechanism by which kinesin motors are thought to pull a membrane nanotube is sketched in Fig. 4.21. First, motors distributed along the nanotube bind to and unbind from the microtubule at rates k_b and k_u, respectively, and the bound motors move with a speed V_0. The values of V_0 and k_u are based on single-motor properties in the absence of an external load. Typical parameter values are $k_b = 5\,\mathrm{s}^{-1}$, $k_u = 0.5\,\mathrm{s}^{-1}$, and $V_0 = 0.6\,\mu\mathrm{m\,s}^{-1}$ [649]. Each motor also feels a drag force due to its motion relative to the membrane nanotube, with the latter moving at speed V. Although the drag on a single motor is small, the cumulative effect on all of the motors can be significant. Motors at the tip of the nanotube exert a normal force that can pull the nanotube out from the vesicle. However, since the speed of kinesin motors decreases with applied load, the motors at the tip move more slowly than those moving along the tube, resulting in an accumulation of motors at the tip which provide the necessary force to generate and sustain the tube.

We now describe a dynamical model of motor-assisted tube extraction due to Leduc et al. [92, 377]. It is first necessary to state a few results regarding the

mechanical properties of vesicular membranes, in order to determine the applied force F required to pull a tube of length L from a vesicle. (For a detailed description of cell mechanics, see [53].) The extraction of a single tube from a spherical vesicle of area A_0 means that the area increases by some amount δA_0, while the volume remains constant. This then generates an increase in membrane tension τ. The excess area $\delta A_0/A_0 = rL/2R_0^2$ where r is the radius of the tube and R_0 is the initial radius of the sphere (typically $R_0 \approx 10\,\mu\text{m}$). For a given membrane tension τ and bending rigidity κ, one finds that [509] (see Ex. 4.8)

$$F = f(\tau) \equiv 2\pi\sqrt{2\kappa\tau}, \quad r = r(\tau) \equiv \sqrt{\kappa/2\tau}. \tag{4.4.18}$$

The relationship between F and L can now be determined by specifying how the membrane tension varies with the excess area. There are two distinct regimes. At low vesicle tensions, τ depends exponentially on the excess area, reflecting the thermodynamics of membrane fluctuations, which leads to the following force-displacement relation:

$$\frac{L}{L_c} = \frac{2F}{F_0} \ln\left(\frac{F}{F_0}\right), \tag{4.4.19}$$

where $L_c = k_B T R_0^2/4\pi\kappa r_0$ is the characteristic length scale at which the increase in F becomes significant, $r_0 = r(\tau_0)$, and $F_0 = f(\tau_0)$. For the given experimental conditions, the force increase was noticeable at tube lengths greater than about $30\,\mu\text{m}$ [377]. If the vesicle is initially already under sufficient tension ($\tau_0 \sim 10^{-4}\,\text{Nm}^{-1}$), then membrane stretching dominates (elastic regime) and the vesicle tension increases linearly with the excess area. The force–displacement relation now takes the form

$$\frac{L}{L_e} = \frac{F}{F_0}\left[\left(\frac{F}{F_0}\right)^2 - 1\right], \tag{4.4.20}$$

where $L_e = R_0^2 F_0^3/(8\pi^3\kappa^2 K_a)$ is the characteristic length in the elastic regime (with K_a the elastic rigidity).

Given the force–displacement function $F = F(L)$, one can now derive dynamical equations for the extraction of a nanotube based on the *force-balance* equation (see also Sects. 8.2 and 8.3):

$$F_M - F(L) + F_D = 0, \tag{4.4.21}$$

where F_M is the force exerted by motors at the tip and F_D is the net drag force on the tube. All the forces must balance, since inertial terms can be neglected in the low Reynolds number regime (see also Box 5B). It turns out that the dominant contribution to the drag force is the accumulative effect of drag due to each of the motors moving with velocity $V_0 - L'$ relative to the membrane, where $L' = dL/dt$. Thus

$$F_D = \xi L(V_0 - L'), \quad \xi = \rho_b \frac{k_B T}{D},$$

where ξ is the motor friction coefficient per unit length, D is the diffusivity, ρ_b is the density of bound motors along the tube, and we have used the Einstein relation.

One finds that friction only becomes significant in regimes where $|L'| \gg V_0$, for which the force-balance equation becomes

$$\xi L \frac{dL}{dt} = F_M - F(L). \tag{4.4.22}$$

The motor force F_M arises from the collective action of the cluster of bound motors at the tip. If n_b is the number of motors at the tip and each independently applies a force f_m, then $F_M = n_b f_m$. The motors at the tip move with a load-dependent velocity V. Assuming a linear force–velocity relation, we have

$$V = V_0 \left(1 - \frac{F_M}{n_b f_s} \right),$$

with f_s the stall force of an individual motor. Finally, setting $L' = V$, the force balance equation (4.4.22) becomes

$$\xi L L' = (1 - V/V_0) n_b f_s - F(L) = (1 - L'/V_0) n_b f_s - F(L),$$

which can be rearranged to give

$$\frac{dL}{dt} = V_0 \frac{n_b - F(L)/f_s}{n_b + \xi V_0 L / f_s}. \tag{4.4.23}$$

Equation (4.4.23) implies that there are two limiting cases. When the number n_b of motors at the tip is sufficiently large, the force on each motor is negligible so that it moves at speed V_0 and $L' \to V_0$. On the other hand, when the motors are not able to sustain a force ($n_b \to 0$), the tube retracts according to

$$\frac{dL}{dt} = -\frac{F(L)}{\xi L}.$$

In order to complete the dynamical description of the system, it is necessary to take into account the dynamics of the bound motors at the tip, that is, the time dependence of n_b. The latter satisfies the conservation equation

$$\frac{dn_b}{dt} = J_b - k_u n_b, \tag{4.4.24}$$

where k_u is the rate of unbinding from the microtubule and J_b is the flux of motors entering the tip region (expressed in the reference frame of the tube); the rate at which tip motors rebind the microtubule can be neglected. Using Kramers rate theory (Sect. 3.3), the unbinding rate of a motor can be related to the force f_m it exerts according to

$$k_u = k_0 \mathrm{e}^{f_m a / k_B T} = k_0 \exp\left([1 - L'/V_0] \frac{f_s a}{k_B T} \right), \tag{4.4.25}$$

where a is a characteristic length scale of the width of the energy barrier between bound and unbound states ($a \approx 1$ nm). The flux J_b is given by

$$J_b = (V_0 - L')\rho_b,$$

which depends on the motor density per unit length ρ_b and the latter is assumed to be uniform for simplicity. Assuming that the rates of binding and unbinding are relatively fast, ρ_b can be related to the total density ρ_0 of motors coating unit area of membrane according to the equilibrium condition

$$\rho_b = 2\pi r \rho_0 \frac{k_b}{k_b + k_0}.$$

Using Eq. (4.4.18) to relate r to the tube force $F(L)$, we have

$$\rho_b = \frac{4\pi^2 \kappa \rho_0}{F(L)} \frac{k_b}{k_b + k_0}. \tag{4.4.26}$$

Finally, substituting for L' in Eq. (4.4.25) and using $\xi = \rho_b k_B T/D$ gives

$$k_u = k_0 \exp\left(\frac{\rho_b k_B T (V_0 L/D) + F(L)}{f_s n_b + \rho_b k_B T (V_0 L/D)} \frac{f_s a}{k_B T} \right). \tag{4.4.27}$$

Following [92], it is convenient to nondimensionalize Eqs. (4.4.23) and (4.4.24) and to reexpress them in terms of the variables (n_b, F) rather than (n_b, L); the force–displacement relations are invertible. First, introduce the so-called processivity length $l_p = V_0/k_0$, where k_0 is the load-free unbinding rate, and then perform the rescalings

$$L \to L/l_p, \quad F \to F/F_0, \quad n_b \to n_b f_s/F_0, \quad t \to k_0 t.$$

In terms of the nondimensionalized variables, the force–displacement relation is written as $L = \beta g(F)$ with $\beta \equiv L_{c,e}/l_p$. Thus, in the thermodynamic regime $g(F) = 2F\ln(F)$ and in the elastic regime $g(F) = F(F^2 - 1)$. Moreover, Eq. (4.4.27) may be expressed in the nondimensional form

$$k_u = \exp\left(\frac{\gamma g(F)/F + F}{n_b + \gamma g(F)/F} \frac{f_s a}{k_B T} \right), \tag{4.4.28}$$

with

$$\gamma = \beta \left(\frac{k_B T V_0}{D f_s} \right) \gamma_0, \quad \gamma_0 = \frac{4\pi^2 \kappa f_s l_p}{F_0^2} \frac{k_b}{k_b + k_0} \rho_0.$$

Equations (4.4.24) and (4.4.23) now become

$$\frac{dn_b}{dt} = \frac{\gamma_0}{F}\frac{F+\gamma g(F)/F}{n_b+\gamma g(F)/F} - \exp\left(\frac{\gamma g(F)/F+F}{n_b+\gamma g(F)/F}\frac{f_s a}{k_B T}\right) n_b \tag{4.4.29a}$$

$$\frac{dF}{dt} = A(F)\frac{n_b - F}{n_b+\gamma g(F)/F}, \tag{4.4.29b}$$

with

$$A(F) = \left[\beta \frac{dg(F)}{dF}\right]^{-1}.$$

Equations (4.4.29) have a single fixed point $(\bar{n}_b, \bar{F})$ with $\bar{n}_b = \bar{F}$ and

$$\frac{\gamma_0}{\bar{F}} = \mathrm{e}^{f_s a/k_B T}\bar{F},$$

that is,

$$\bar{F} = \sqrt{\gamma_0 \mathrm{e}^{-f_s a/k_B T}}.$$

Since the tube length $L \geq 0$ and when $L = 0$ the membrane tension $\tau = \tau_0$ and $F = F_0$ (in physical variables), it follows that $F(L) \geq F_0$ and thus $\bar{F} \geq 1$ (in dimensionless variables). This implies that a nanotube can only be extracted if $\gamma_0 \mathrm{e}^{-f_s a/k_B T} > 1$, so that the density of coating ρ_0 has to exceed a minimum value $\rho_{\min}$ given by

$$\rho_{\min} = \frac{F_0^2}{4\pi^2 \kappa l_p f_s \mathrm{e}^{-f_s a/k_B T}}\frac{k_b + k_0}{k_b}.$$

One can now analyze the stability of the fixed point (assuming it exists), by linearizing equation (4.4.29) about the fixed point and determining the eigenvalues of the resulting Jacobian. There are four dimensionless parameters in Eq. (4.4.24), namely, β, γ, γ_0, and $f \equiv f_s a/k_B T$. It is convenient to consider the equivalent set of parameters $p \equiv \rho_m/\rho_0, f, \beta$ and $\xi_0 \equiv \gamma\beta/\gamma_0 = k_B T V_0/(D f_s)$. Fixing ξ_0, one finds that the fixed point is stable provided that $f < 1$, whereas if $f > 1$, then there are biophysically realistic parameter regimes in which the fixed point undergoes a Hopf bifurcation along the lines outlined in Box 4B [92, 377], resulting in the occurrence of a stable periodic orbit or limit cycle as observed experimentally. This is illustrated in Fig. 4.22, which shows a bifurcation diagram in the (p, β)-plane for $f = 2.03$ and $\xi_0 = 4.1 \times 10^{-4}$ [92]. A region of oscillatory behavior occurs, consistent with the hypothesis that the collective behavior of a cluster of molecular motors is pulling the nanotube at the tip. As the tube grows, the membrane tension increases so that the tip motors have to exert a stronger force in order to maintain elongation of the tube. However, once the tension is too high, the incoming flux of motors to the tip is not sufficient to generate a large enough cluster to generate the necessary force, and the tube starts to shrink. This relaxation allows the reformation of a large enough cluster to start growing the nanotube again.

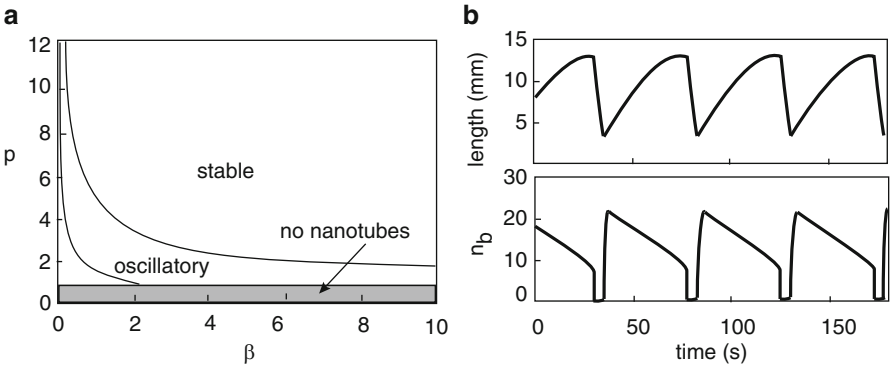


Fig. 4.22: Spontaneous oscillations in the model of nanotube extraction given by Eq. (4.4.29). (**a**) Bifurcation diagram showing different dynamical regimes in the (β, p)-plane for $f = 2.03$ and $\xi_0 = 4.1 \times 10^{-4}$. (**b**) Sketch of oscillations in the tube length and the number n_b of bound motors at the tip for $\rho_{\min}/\rho_0 = 7.3$, $f = 2.03$ and $\xi_0 = 4.1 \times 10^{-4}$ and $\beta = 1.73$. (Redrawn from [92].)

Box 4B. The Hopf bifurcation.

Any differential equation describing the dynamics of some biological system will depend on one or more parameters μ. This can be made explicit by writing

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}(t); \mu), \quad \mathbf{x} \in \mathbb{R}^n, \mu \in \mathbb{R}^m.$$

For simplicity, suppose that only one parameter is varied and set $m = 1$. The dynamical system is said to have a *bifurcation* at the critical value $\mu = \mu_c$ if there is a change in the (topological) structure of trajectories as μ crosses μ_c. (For a detailed introduction to bifurcation theory see the book *Elements of Applied Bifurcation Theory* by Kuznetsov [366].) In the case of local bifurcations, this means that there is a change in the number and/or stability of equilibria or fixed points. A *Hopf bifurcation* occurs when a fixed point changes stability as μ crosses μ_c resulting in the emergence of a small amplitude limit cycle. If the limit cycle is stable and surrounds an unstable fixed point, then the Hopf bifurcation is said to be *supercritical*, whereas if the limit cycle is unstable and surrounds a stable fixed point, then it is said to be *subcritical*. Before stating the general conditions for the occurrence of a Hopf bifurcation in a planar system ($n = 2$), it is useful to consider an explicit example:

$$\frac{dx}{dt} = \mu x + \omega y + \sigma x(x^2 + y^2), \quad \frac{dy}{dt} = -\omega x + \mu y + \sigma y(x^2 + y^2), \quad (4.4.30)$$

with $\sigma = \pm 1$. This pair of equations has a single fixed point at the origin. Transforming to polar coordinates by setting $x = r\cos\theta, y = r\sin\theta$, we have

$$\frac{dr}{dt} = r(\mu + \sigma r^2), \quad \frac{d\theta}{dt} = -\omega,$$

with $r \geq 0$. It immediately follows that all nonzero solutions rotate clockwise at the same angular frequency ω. Suppose that $\sigma = -1$. If $\mu \leq 0$, then $\dot{r} < 0$ for all $r > 0$ and trajectories converge to the fixed point at the origin, which is stable. On the other hand, if $\mu > 0$ then $\dot{r} < 0$ for $r \in (\sqrt{\mu}, \infty)$ and $\dot{r} > 0$ for $r \in (0, \sqrt{\mu})$. Hence, the origin is now an unstable fixed point, whereas there is a stable periodic orbit at $r = \sqrt{\mu}$ (see Fig. 4.23). In other words, the system undergoes a supercritical Hopf bifurcation at the critical value $\mu = \mu_c = 0$. Similarly, if $\sigma = +1$ then the fixed point undergoes a subcritical Hopf bifurcation with an unstable limit cycle existing for $\mu < 0$.

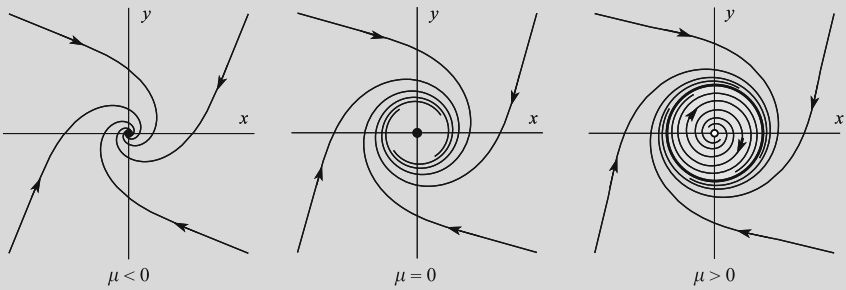


Fig. 4.23: Supercritical Hopf bifurcation. Phase portraits for system $\dot{x} = \mu x + y - x(x^2 + y^2)$ and $\dot{y} = -x + \mu y - y(x^2 + y^2)$ for the three cases $\mu < 0$, $\mu = 0$, and $\mu > 0$

Hopf bifurcation theorem (2D). Consider the planar dynamical system

$$\frac{dx}{dt} = f(x, y; \mu), \quad \dot{y} = g(x, y; \mu)$$

for the single parameter μ. Suppose that there exists a fixed point at $(x, y) = (0, 0)$, say. Linearizing about the origin gives the linear system

$$\frac{d}{dt}\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{A}(\mu)\begin{pmatrix} x \\ y \end{pmatrix},$$

where $\mathbf{A}(\mu)$ is the Jacobian

$$\mathbf{A}(\mu) = \begin{pmatrix} \partial_x f(0,0;\mu) & \partial_y f(0,0;\mu) \\ \partial_x g(0,0;\mu) & \partial_y g(0,0;\mu) \end{pmatrix}.$$

Let the eigenvalues of the Jacobian be given by the complex conjugate pair $\lambda(\mu), \overline{\lambda}(\mu) = \alpha(\mu) \pm i\beta(\mu)$. Furthermore, suppose that at a certain value $\mu = \mu_c$ (with $\mu_c = 0$, say), the following conditions hold:

1. $\alpha(0) = 0$, $\beta(0) = \omega \neq 0$ with $\text{sign}(\omega) = \text{sign}[\partial_\mu g(0,0;0)]$.
2. $\left.\dfrac{d\alpha(\mu)}{d\mu}\right|_{\mu=0} \neq 0$ (transversality condition) .
3. The so-called first Liapunov coefficient l_1 doesn't vanish, $l_1 \neq 0$ (genericity condition), where

$$l_1 = \frac{1}{16}(f_{xxx} + f_{xyy} + g_{xxy} + g_{yyy}) + \frac{1}{16\omega}(f_{xy}[f_{xx} + f_{yy}] - g_{xy}[g_{xx} + g_{yy}] - f_{xx}g_{xx} + f_{yy}g_{yy}),$$

with $f_{xy} = \partial_x \partial_y f(0,0;0)$ etc.

Then a unique curve of periodic solutions bifurcates from the origin as μ crosses zero. The amplitude of the limit cycle grows like $\sqrt{|\mu|}$ and the period tends to $2\pi/\omega$ as $|\mu| \to 0$. Suppose for the sake of illustration that $\alpha'(\mu) > 0$ and set $\sigma = \text{sign}(l_1)$.

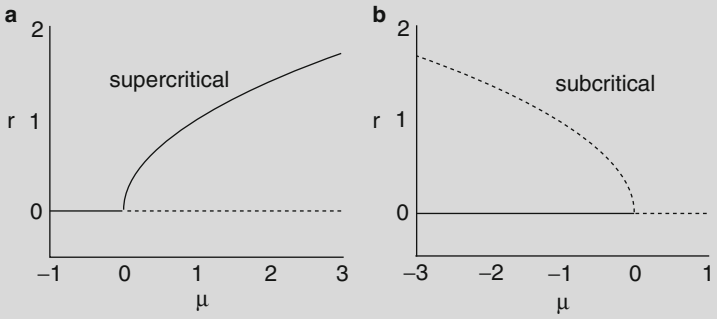


Fig. 4.24: Bifurcation diagrams for (**a**) supercritical ($\sigma = -1$) and (**b**) subcritical ($\sigma = +1$) Hopf bifurcations in the same dynamical system as Fig. 4.23

If $\sigma = -1$, then the fixed point is asymptotically stable for $\mu \leq 0$ and unstable for $\mu > 0$. Moreover, there is a unique and stable periodic orbit that exists for $\mu > 0$, corresponding to the case of a supercritical Hopf bifurcation (see Fig. 4.24a).

If $\sigma = +1$, then the fixed point is asymptotically stable for $\mu < 0$ and unstable for $\mu \geq 0$. Moreover, an unstable limit cycle exists for $\mu < 0$, and we have a subcritical Hopf bifurcation (see Fig. 4.24b).

Finally, sufficiently close to the bifurcation point, the dynamical system is locally topologically equivalent to the *normal form* given by Eq. (4.4.30).

There is also an n-dimensional version of the Hopf bifurcation theorem for $n \geq 3$. In particular, suppose that the $n \times n$ Jacobian matrix has a pair of complex conjugate eigenvalues $\lambda(\mu), \overline{\lambda}(\mu) = \alpha(\mu) \pm i\beta(\mu)$ and all other eigenvalues have negative real parts, $\mathrm{Re}[\lambda_j(\mu)] < 0$ for $j = 3, \ldots, n$. It can then be shown that the system converges to a family of smooth two-dimensional invariant manifolds W_c (trajectories starting in W_c stay within W_c) and the above Hopf bifurcation theorem then applies to the effective dynamics on W_c (Fig. 4.25).

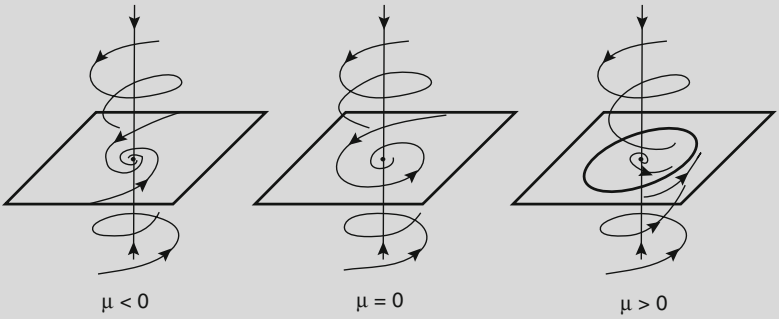


Fig. 4.25: Hopf bifurcation in 3D

Linear stability analysis [221]. In the statement of the Hopf bifurcation theorem, the (local) stability of a fixed point was determined in terms of the eigenvalues of the Jacobian. We now explain this in more detail. Let us return to a planar dynamical system written as

$$\frac{dx_1}{dt} = f_1(x_1, x_2), \quad \frac{dx_2}{dt} = f_1(x_1, x_2),$$

where we have suppressed the bifurcation parameter. Introduce the vector notation $\mathbf{x} = (x_1, x_2)^T$, $\mathbf{f} = (f_1, f_2)^T$, and suppose that $\mathbf{x} = \mathbf{x}^*$ is a fixed point for which $\mathbf{f}(\mathbf{x}) = \mathbf{0}$. Set $\mathbf{y} = \mathbf{x} - \mathbf{x}^*$ and Taylor expand the ODEs to first order in $\mathbf{y}$:

$$\frac{dy_j}{dt} = \frac{dx_j}{dt} = f_j(x_1^* + y_1, x_2^* + y_2) = f_j(x_1^*, x_2^*) + \sum_{k=1,2} \frac{\partial f_j}{\partial x_k}(x_1^*, x_2^*) y_k + \text{ h. o. t.}.$$

Imposing the fixed point conditions and dropping the higher-order terms then yields the linear equation

$$\frac{d\mathbf{y}}{dt} = \mathbf{A}\mathbf{y}, \quad A_{jk} = \frac{\partial f_j}{\partial x_k}(x_1^*, x_2^*). \tag{4.4.31}$$

Trying a solution of the form $\mathbf{y} = \mathbf{v}e^{\lambda t}$ generates the eigenvalue equation

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}.$$

This will have a nontrivial solution provided that $\mathbf{A} - \lambda \mathbf{I}$ is non-invertible, where $\mathbf{I}$ is the 2×2 unit matrix, which means that the pair of eigenvalues is a solutions to the characteristic equation

$$\text{Det}[\mathbf{A} - \lambda \mathbf{I}] = 0. \tag{4.4.32}$$

This yields a quadratic equation for λ, which can be factorized as

$$(\lambda - \lambda_1)(\lambda - \lambda_2) = 0,$$

where the roots $\lambda_{1,2}$ are the eigenvalues. If these eigenvalues are distinct, then the general solution to the linear ODE can be written as

$$\mathbf{y}(t) = \sum_{j=1,2} c_j \mathbf{v}_j e^{\lambda_j t}$$

for constant coefficients c_j, where $\mathbf{v}_j$ is the unit eigenvector corresponding to λ_j.

Expanding the factorized equation and using the fact that $\text{Tr}[\mathbf{A}] = \lambda_1 + \lambda_2$ and $\text{Det}[\mathbf{A}] = \lambda_1 \lambda_2$ gives

$$\lambda^2 - \text{Tr}[\mathbf{A}]\lambda + \text{Det}[\mathbf{A}] = 0.$$

Thus

$$\lambda_{1,2} = \frac{1}{2}\left[\text{Tr}[\mathbf{A}] \pm \sqrt{\text{Tr}[\mathbf{A}]^2 - 4\text{Det}[\mathbf{A}]}\right]. \tag{4.4.33}$$

It follows from the above analysis that the fixed point $\mathbf{x}^*$ will be stable provided that $\text{Re}[\lambda_{1,2}] < 0$, since the perturbations $\mathbf{y}_j(t) = \mathbf{v}_j e^{\lambda_j t} \to 0$ as $t \to \infty$. Using Eq. (4.4.33), the condition for linear stability is $\text{Tr}[\mathbf{A}] < 0$. Moreover, reintroducing the bifurcation parameter μ, the first two conditions for a Hopf bifurcation at $\mu = \mu_c$ can be expressed as

$$\text{Tr}[\mathbf{A}(\mu_c)] = 0, \quad \text{Det}[\mathbf{A}(\mu_c)] > 0. \tag{4.4.34}$$

Then $\lambda_{1,2}(\mu_c) = \pm i\omega_0$ with $\omega_0 = \sqrt{\text{Det}[\mathbf{A}(\mu_c)]}$.

4.4.4 Rigidly Linked Molecular Motors

As our final example of a motor assembly, consider the case where N motors are rigidly coupled to a common backbone, which is connected to a fixed cytoskeletal structure via a spring K [307, 308] (see Fig. 4.26). Each motor is treated as a two-state Brownian ratchet, such that the motor sees an asymmetric, l-periodic potential $V_1(x)$ when bound to a cytoskeletal filament and a flat potential V_2 when unbound. The motors switch between the two states with position-dependent transition rates $\omega_{1,2}(x)$, which do not satisfy detailed balance due to ATP hydrolysis. Cooperativity arises due to the global motion of the backbone relative to the filament, which simultaneously modifies the positions of all the motors. Such a configuration mimics the experimental protocol known as a *motility assay* as well as the coupling of myosin motor cross bridges in muscles.

Suppose that the displacement of the backbone at time t is $Y(t)$ and the center of masses of the motors is separated by a uniform spacing q on the backbone. It follows that the position of the nth motor at time t is $x_n(t) = Y(t) + nq$. Each motor is either in the bound state ($\sigma = 1$) or the unbound state ($\sigma = 2$) and the energy of the nth motor in state σ is $V_\sigma(x_n(t))$. However, since $V_\sigma(x+l) = V_\sigma(x)$ for all x, we need only specify the position of the motor using the cyclic coordinate $\xi_n(t) = x_n(t)$ mod l with $0 < \xi_n < l$. Let $P_\sigma(\xi,t)$ be the probability that there exists a motor in state σ at cyclic position ξ at time t. It follows that

$$P(\xi,t) \equiv P_1(\xi,t) + P_2(\xi,t) = \frac{1}{N}\sum_{n=1}^{N} \delta(\xi - \xi_n(t)).$$

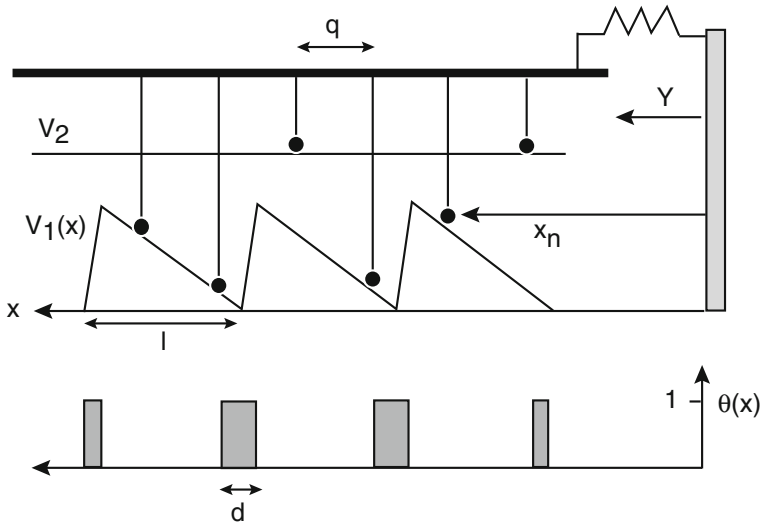


Fig. 4.26: Ensemble of molecular motors coupled rigidly to a moving backbone. See text for details

A major simplification occurs if the periodic spacing q of the motors is incommensurate with the period l of the filament track, that is, l/q is an irrational number. In the limit $N \to \infty$, the motor positions ξ_n form a dense subset on $[0,l]$ such that $P(x,t) \to 1/l$. The equations of motion for P_σ are [307]

$$\frac{\partial P_1}{\partial t} + v\frac{\partial P_1}{\partial \xi} = -\omega_1 P_1 + \omega_2 P_2 \tag{4.4.35a}$$

$$\frac{\partial P_2}{\partial t} + v\frac{\partial P_2}{\partial \xi} = \omega_1 P_1 - \omega_2 P_2, \tag{4.4.35b}$$

where $v = dY/dt$ is the velocity of the backbone and the transition rates satisfy Eq. (4.3.32), that is,

$$\Omega(\xi) = \omega_1(\xi) - \omega_2(\xi)\mathrm{e}^{[V_1(\xi))-V_2]/k_BT}.$$

It is convenient to set $\Omega(\xi) = \Omega\theta(\xi)$ and $\int_0^l \theta(\xi)d\xi = 1$. For concreteness, ω_2 is a constant and $\omega_1(\xi)$ is determined by taking $\theta(\xi)$ to be a periodic sequence of square pulses of width d (see Fig. 4.26). Finally, the velocity v is determined by the force-balance equation

$$F_{\text{ext}} = \mu v(t) + KY(t) + \int_0^l (P_1\partial_\xi V_1 + P_2\partial_\xi V_2)d\xi, \tag{4.4.36}$$

where F_{ext} is an externally applied force (per motor). The first term on the right-hand side is a frictional force with damping coefficient μ, the second term is the elastic force from the spring with elastic modulus NK, and the third term is the force due to the potentials.

For an incommensurate system with $P_2 = l^{-1} - P_1$, and a soft spring ($K = 0$), there exists a steady state that satisfies the pair of equations

$$v\frac{\partial P_1}{\partial \xi} = -(\omega_1 + \omega_2)P_1 + \frac{\omega_2}{l} \tag{4.4.37}$$

and (since V_2 is constant)

$$F_{\text{ext}} = \mu v + \int_0^l P_1\partial_\xi V_1 d\xi. \tag{4.4.38}$$

Let us nondimensionalize the system by fixing the length and time units such that $l, q, \omega_2 = O(1)$ and suppose that $v \ll 1$. The solution for P_1 may then be expanded as a Taylor series in v:

$$P_1(\xi) = \sum_{n=0}^{\infty} v^n P_1^{(n)}(\xi).$$

Substituting into Eq. (4.4.37) gives

$$P_1^{(0)}(\xi) = \frac{1}{l}\frac{\omega_2(\xi)}{\omega_1(\xi)+\omega_2(\xi)}, \quad P_1^{(n)}(\xi) = -\frac{1}{\omega_1(\xi)+\omega_2(\xi)}\partial_\xi P_1^{(n-1)}(\xi). \tag{4.4.39}$$

Substituting the Taylor expansion of P_1 into Eq. (4.4.38) then yields

$$F_{\rm ext} - F^{(0)} = (\mu + F^{(1)})v + \sum_{n=2}^{\infty} v^n F^{(n)}, \text{ for } F^{(n)} = \int_0^l P_1^{(n)} \partial_\xi V_1 d\xi. \tag{4.4.40}$$

It follows from Eq. (4.3.32) that

$$P_1^{(0)}(\xi) = l^{-1} \frac{1}{{\rm e}^{(V_1(\xi)-V_2)/k_BT} + 1 + \Omega\theta(\xi)/\omega_2(\xi)}.$$

Hence, if detailed balance holds ($\Omega = 0$), then from the periodicity of the potentials,

$$F^{(0)} = -\frac{k_BT}{l} \int_0^l \partial_\xi \ln\left[1 + {\rm e}^{-(V_1(\xi)-V_2)/K_BT}\right] d\xi = 0.$$

One also finds that $F^{(1)} > 0$, that is, there is an effective increase in friction so $v = 0$ in the absence of an external force. On the other hand, when detailed balance is broken, two important features arise [307]:

(i) If $\Omega \neq 0$ and the potential $V_1(\xi)$ is asymmetric, then $F^{(0)} \neq 0$ and the system can do work against an external load as previously found for a single motor. One subtle point is that we still have $F^{(0)} = 0$ if the transition rates are homogeneous, which differs from the single-motor case.

(ii) If $\Omega \neq 0$, then $F^{(1)}$ may become negative as an emergent feature of cooperativity. This can overcome the effects of external damping, resulting in a nonzero velocity in the absence of an external force. Note that in contrast to a single motor, it is no longer necessary for asymmetric potentials or transition rates.

In order to explore feature (ii) in more detail, suppose that both the potential $V_1(\xi)$ and $\theta(\xi)$ are symmetric functions, with $\theta(\xi)$ determining the deviation from detailed balance. The perturbation analysis implies that $P^{(n)}(\xi)$ is symmetric for even n and antisymmetric for odd n [see Eq. (4.4.39)]. Since taking a derivative converts a symmetric function to an antisymmetric one and vice versa, we see from Eq. (4.4.40) that $F^{(n)} = 0$ for all even n. Therefore, we have

$$F_{\rm ext} = (\mu + F^{(1)})v + v^3 F^{(3)} + O(v^5). \tag{4.4.41}$$

This represents the normal form of an imperfect pitchfork bifurcation. First consider the case $F_{\rm ext} = 0$. Suppose that $F^{(1)}$ is a monotonically decreasing function of Ω such that $\mu + F^{(1)} = 0$ at a critical value $\Omega = \Omega_c$. Then for $\Omega < \Omega_c$, the only solution is $v = 0$ and the system does not move. However, at the critical point $\Omega = \Omega_c$, the zero velocity state loses stability and two new stable solutions $v_\pm$ emerge via spontaneous symmetry breaking, with $v_\pm \sim \sqrt{\Omega - \Omega_c}$ close to the bifurcation point (see Fig. 4.27a). One can also determine the relationship between an external force and velocity for different values of Ω as illustrated in Fig. 4.27b.

Now suppose that the spring constant K of the spring connecting the backbone to the cytoskeleton is nonzero. It is now possible for the system to exhibit spontaneous oscillations [309]. In order to show this, consider the linear stability of the zero

velocity steady-state solution of Eqs. (4.4.35) and (4.4.36) of zero external force. The steady-state solution is given by

$$P_1(\xi) = R(\xi) \equiv \frac{1}{l}\frac{\omega_2(\xi)}{\omega_1(\xi)+\omega_2(\xi)}, \quad Y = Y_0 \equiv -\frac{1}{K}\int_0^l R(\xi)\partial_\xi V_1(\xi)d\xi.$$

Linearize about the steady-state solution by setting

$$P_1(\xi,t) = R(\xi) + p(\xi)e^{\lambda t}, \quad Y(t) = Y_0 + ye^{\lambda t},$$

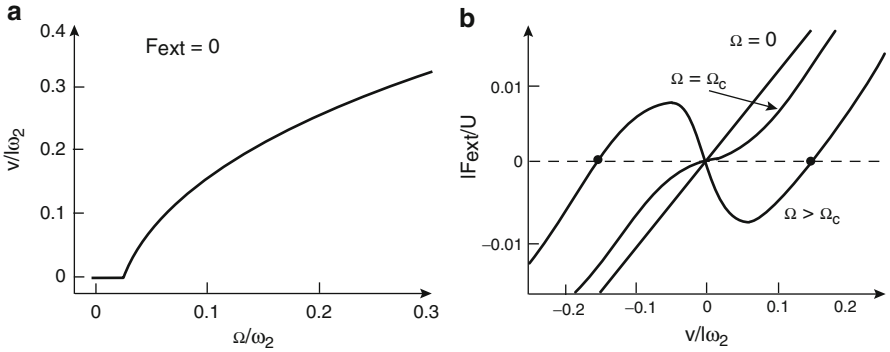


Fig. 4.27: Ensemble of molecular motors coupled rigidly to a moving backbone. (**a**) There exists a critical amplitude $\Omega = \Omega_c$, where Ω measures the size of the deviation from detailed balance, beyond which the motor assembly has a nonzero velocity in the absence of an external force. (**b**) Sketch of force–velocity curves for different amplitudes Ω and a symmetric potential $V_1(x)$ of height U and periodicity l, illustrating spontaneous symmetry breaking. For the dimensionless parameter values $d/l = 0.1$ and $\mu\omega_2 l^2/U = 0.1$ one finds that $\Omega_c/\omega_2 \approx 0.026$. (Redrawn from Julicher and Prost [307].)

and $v(t) = \partial_t Y(t) = \lambda y e^{\lambda t}$, we find

$$\lambda p(\xi) + \lambda y R'(\xi) = -(\omega_1(\xi)+\omega_2(\xi))p(\xi)$$
$$0 = \mu\lambda y + Ky + \int_0^l p(\xi)V_1'(\xi)d\xi.$$

Expressing $p(\xi)$ in terms of y and eliminating y yields the eigenvalue equation

$$\mu + \frac{K}{\lambda} = \int_0^l \frac{R'(\xi)V_1'(\xi)}{\lambda+\omega_1(\xi)+\omega_2(\xi)}d\xi. \tag{4.4.42}$$

The zero velocity state is stable if $\mathrm{Re}\,\lambda < 0$. An instability will occur if there exists a critical value $\Omega = \Omega_c(K)$ for which $\lambda = i\omega$. Note that if $K = 0$, then the left-hand side is real, which implies that λ is real and destabilization of the zero velocity

state cannot occur via a Hopf bifurcation. On the other hand, if $K > 0$, then a Hopf bifurcation can occur resulting in a limit cycle oscillation with frequency $\omega(K)$ (see Box 4B). An amplitude equation for spontaneous oscillations close to the bifurcation point can be obtained using Fourier analysis and perturbation theory [308] (see Ex. 4.9).

4.5 Appendix: Statistical Mechanics of Polymers

In the bulk of the book, we treat F-actin and microtubules as rigid or semirigid polymer filaments. These provide one-dimensional tracks for active intracellular transport by molecular motors and self-organize through polymerization into the various structures of the cytoskeleton responsible for cell mitosis and cell motility (Chap. 8). We will neglect statistical mechanical effects in these applications. On the other hand, biopolymers such as DNA and proteins are much more flexible relative to their lengths, which means one has to take into account the wide range of different configurations that can occur through bending and folding of the polymers, as well as the associated energetics. The statistical mechanics of polymers will be important when considering translocation of biopolymers through membrane pores (Sect. 7.3). The examples in this section also help illustrate some of the basic principles of statistical mechanics.

Example 4.1 (Random walk model of a polymer (1D)). Consider a simple 1D model of a flexible polymer. The polymer is represented as a sequence of links of length a that either point in the positive or negative x-direction with equal probability. One can thus treat a given configuration or microstate of the polymer as a sample trajectory of an unbiased random walk on a 1D lattice with spacing a (see Ex. 4.10). Suppose that there are N links in the chain and one end of the chain is fixed at the origin. Let n denote the number of links pointing in the positive x-direction. It follows that the other end of the chain (end-to-end distance) is at $x = (2n - N)a$. We will assume that N is sufficiently large so that x and n can be treated as continuous variables. If we ignore any energy contributions from the elastic stretching, bending, or twisting of the polymer, then the energy Φ of any configuration is zero. However, stretching the polymer by pulling on the free end at x is resisted by an entropic force. In order to show this, we note that the number of configurations or internal microstates for fixed n or x is given by the combinatorial factor (see also Sect. 1.4)

$$\Omega(n) = \frac{N!}{n!(N-n)!}. \tag{4.5.1}$$

Taking logs and using Stirling's formula (2.1.5) we have the entropy

$$S(n) = k_B[N\ln N - n\ln n - (N-n)\ln(N-n)], \tag{4.5.2}$$

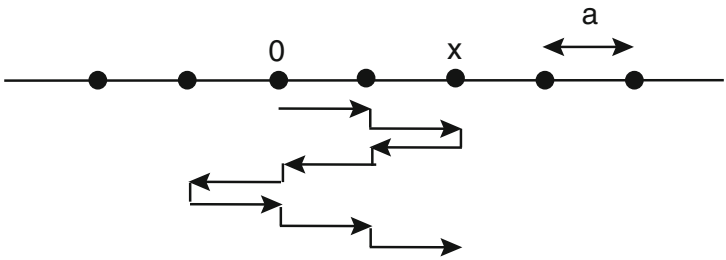


Fig. 4.28: Random walk model of a 1D polymer. The links have been displaced in the vertical direction for illustrative purposes

which can be reexpressed in terms of x according to

$$S(x) = k_B \left(N \ln N - \frac{x+Na}{2a} \ln \left[\frac{x+Na}{2a} \right] - \frac{Na-x}{2a} \ln \left[\frac{Na-x}{2a} \right] \right). \tag{4.5.3}$$

The entropic force is then

$$f_{\text{ent}}(x) = T \frac{dS(x)}{dx} = -\frac{k_B T}{2a} \ln \left[\frac{1+x/Na}{1-x/Na} \right]. \tag{4.5.4}$$

Assuming that $x \ll Na$, we can Taylor expand to first order in x to obtain Hooke's law for an elastic spring:

$$f_{\text{ext}}(x) \approx -\frac{k_B T}{Na^2} x. \tag{4.5.5}$$

The minus sign means that the polymer resists stretching, since the force is in the negative x-direction. The origin of the entropic force is that when a polymer is stretched it becomes less random, in the sense that there are less configurations for larger x. In order to maintain the displacement x, the environment has to exert an opposing force $f = -f_{\text{net}}$. Inverting Eq. (4.5.4), we obtain the force–displacement relation

$$z \equiv \frac{x}{Na} = \tanh \left[\frac{fa}{k_B T} \right]. \tag{4.5.6}$$

Example 4.2 (Freely-jointed chain model of a polymer (3D)). Suppose that we represent a polymer in 3D as a chain of N segments; each of which is described by a vector $\mathbf{a}_j$ with $|\mathbf{a}_j| = a$ the length of the segment and the direction of $\mathbf{a}_j$ representing the orientation of the segment (see Fig. 4.29). One end of the polymer is fixed at the origin, so that the end-to-end displacement of the chain is given by $\mathbf{r} = \sum_{j=1}^{N} \mathbf{a}_j$. Suppose that the orientations of the segments are random in the sense that, averaging over a large population of identical polymers, we have $\langle \mathbf{a}_j \cdot \mathbf{a}_i \rangle = 0$ for all $i \neq j$. A fixed force f is applied to the free end of the polymer in the z-direction using an external load, for example. The configuration of the polymer is specified by the spherical polar angles (θ_j, ϕ_j) of the N links, $j = 1, \ldots, N$, with $0 < \theta_j \leq \pi$ and $0 \leq \phi_j \leq 2\pi$. The total energy of the system consisting of the polymer in a given

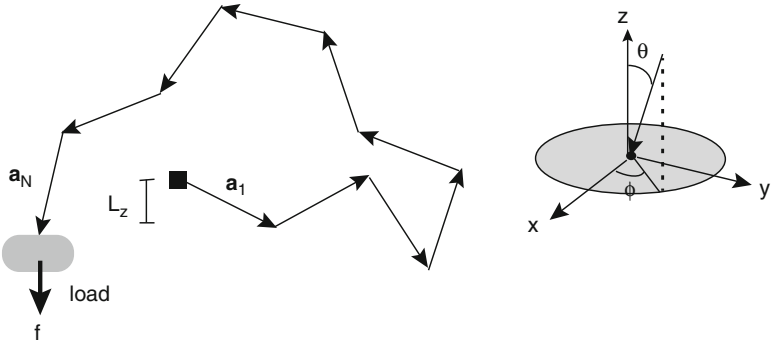


Fig. 4.29: Freely jointed chain model of a 3D polymer. A load maintains a constant force f in the z direction

configuration and the applied load is $E = -\sum_{j=1}^{N} fa\cos(\theta_j)$, which is the work done by the load in moving links in a direction perpendicular to the (x,y)-plane. The Boltzmann–Gibbs distribution with $\Theta = (\theta_1, \ldots, \theta_N)$ and $\Phi = (\phi_1, \ldots, \phi_N)$ is

$$p(\Theta, \Phi) = \frac{1}{Z} \mathrm{e}^{(fa/k_BT)\sum_{j=1}^{N}\cos\theta_j} = \frac{1}{Z}\prod_{j=1}^{N} \mathrm{e}^{(fa/k_BT)\cos\theta_j}, \tag{4.5.7}$$

with the partition function obtained by integrating over all solid angles, $d\Omega = \prod_{j=1}^{N} \sin\theta_j d\theta_j d\phi_j$:

$$Z = \int \mathrm{e}^{(fa/k_BT)\sum_{j=1}^{N}\cos\theta_j} d\Omega \tag{4.5.8}$$

$$= \prod_{j=1}^{N} \left[\int_0^{2\pi} \int_0^{\pi} \mathrm{e}^{(fa/k_BT)\cos\theta_j} \sin\theta_j d\theta_j d\phi_j \right] = Z_1^N, \tag{4.5.9}$$

where

$$\begin{aligned} Z_1 &= \int_0^{2\pi} \int_0^{\pi} \sin\theta \mathrm{e}^{(fa/k_BT)\cos\theta} d\theta d\phi \\ &= 2\pi \int_0^{2\pi} \mathrm{e}^{(fa/k_BT)\cos\theta} d\cos\theta \\ &= 2\pi \int_{-1}^{1} \mathrm{e}^{(fa/k_BT)x} dx = \frac{4\pi k_B T}{fa} \sinh\left(\frac{fa}{k_BT}\right). \end{aligned}$$

It follows from the above analysis that the Boltzmann–Gibbs distribution can be factorized into a product of distributions for the N independent links,

$$p(\Theta, \Phi) = \prod_{j=1}^{N} p_1(\theta_j, \phi_j), \quad p_1(\theta_j, \phi_j) = Z_1^{-1} \mathrm{e}^{(fa/k_BT)\cos\theta_j}. \tag{4.5.10}$$

In order to derive a force–displacement relation, it is necessary to determine the mean displacement of the polymer in the z direction, $L_z = \mathbf{r} \cdot \mathbf{e}_z$:

$$\begin{aligned}\langle L_z \rangle &\equiv \int \left[a \sum_{j=1}^{N} \cos\theta_j \right] p(\Theta, \Phi) d\Omega \\ &= Na \int_0^{2\pi} \int_0^{\pi} \cos\theta \, p_1(\theta, \phi) \sin\theta d\theta d\phi.\end{aligned}$$

The integral can be evaluated by noting that

$$\frac{d \ln Z_1}{df} = \frac{a}{k_B T} \frac{1}{Z_1} \int_0^{2\pi} \int_0^{\pi} \cos\theta \sin\theta e^{(fa/k_B T)\cos\theta} d\theta d\phi,$$

which implies that

$$\langle L_z \rangle = N k_B T \frac{d \ln Z_1}{df} = Na \left[\coth\left(\frac{fa}{k_B T} \right) - \frac{k_B T}{fa} \right]. \tag{4.5.11}$$

In the small force limit, this reduces to Hooke's law:

$$f = k \langle L_z \rangle, \quad k = 3 k_B T / N a^2. \tag{4.5.12}$$

One possible point of confusion is that in Example 4.1, the position x of the polymer was treated as deterministic, whereas in this case L_z is fluctuating. The two pictures are consistent once one notes that for a large polymer (large N) fluctuations in the end-to-end distance are negligible, which is a consequence of the law of large numbers.

Example 4.3 (The Ising model of a polymer). The next level of complexity in the statistical mechanical modeling of polymers is to incorporate elastic effects such as bending, stretching, and twisting. Here we will consider a simplified model, which is equivalent to the classical *Ising model* of magnetic spins [102]. Let us return to the example of a 1D polymer consisting of N links of length a. Denote the state of each link by the binary variable σ_i with $\sigma_i = 1$ ($\sigma = -1$) if the link points in the positive (negative) x-direction. Suppose that an external load maintains a constant force f in the positive x-direction. The total extension of the polymer is $x = a \sum_{j=1}^{N} \sigma_j$. In contrast to the random walk model, suppose that when two neighboring links point in opposite directions, they contribute an extra $2\gamma k_B T$ of energy, where γ is some cooperativity parameter. (This could represent an effective bending energy.) Then the total energy of the polymer plus load for a given configuration $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_N)$ is

$$E[\boldsymbol{\sigma}] = -fa \sum_{j=1}^{N} \sigma_j - \gamma k_B T \sum_{j=1}^{N-1} \sigma_j \sigma_{j+1}. \tag{4.5.13}$$

The corresponding Boltzmann–Gibbs distribution is

$$p(\boldsymbol{\sigma}) = Z^{-1} \mathrm{e}^{-E[\boldsymbol{\sigma}]/k_B T}, \tag{4.5.14}$$

and the partition function is

$$Z = \sum_{\sigma_1 = \pm 1} \cdots \sum_{\sigma_N = \pm 1} \mathrm{e}^{\alpha \Sigma_{j=1}^{N} \sigma_j + \gamma \Sigma_{j=1}^{N-1} \sigma_j \sigma_{j+1}}, \tag{4.5.15}$$

with $\alpha = fa/k_B T$. The partition function can be treated as a generating function for the displacement x, that is,

$$\langle x \rangle = k_B T \frac{d}{df} \ln Z[f] = a \frac{d}{d\alpha} Z[\alpha]. \tag{4.5.16}$$

A well-known result from statistical mechanics is that the 1D Ising model can be solved exactly. In particular, one can derive an exact expression for Z using transfer matrices [102]. First rewrite Z in the more suggestive form

$$Z = \sum_{s_1} \cdots \sum_{s_N} \left[\mathrm{e}^{\alpha(\sigma_1 + \sigma_2)/2 + \gamma \sigma_1 \sigma_2} \right] \left[\mathrm{e}^{\alpha(\sigma_2 + \sigma_3)/2 + \gamma \sigma_2 \sigma_3} \right] \cdots \left[\mathrm{e}^{\alpha(\sigma_N + \sigma_1)/2 + \gamma \sigma_N \sigma_1} \right].$$

We can view each term on the right-hand side as the element of a matrix $\mathbf{T}$ with matrix elements labeled by σ_1, σ_2, etc., that is, $T_{\sigma_1 \sigma_2} = \mathrm{e}^{\alpha(\sigma_1 + \sigma_2)/2 + \gamma \sigma_1 \sigma_2}$. Hence,

$$\mathbf{T} = \begin{pmatrix} T_{11} & T_{1-1} \\ T_{-11} & T_{-1-1} \end{pmatrix} = \begin{pmatrix} \mathrm{e}^{\alpha+\gamma} & \mathrm{e}^{-\gamma} \\ \mathrm{e}^{-\gamma} & \mathrm{e}^{-\alpha+\gamma} \end{pmatrix}.$$

In terms of the transfer matrix $\mathbf{T}$

$$Z = \sum_{\sigma_1} \cdots \sum_{\sigma_N} T_{\sigma_1 \sigma_2} T_{\sigma_2 \sigma_3} \cdots T_{\sigma_N \sigma_1} = \mathrm{Tr}[\mathbf{T}^N],$$

where we have used the standard rules of matrix multiplication. It can then be shown that for large N, $Z \approx \lambda_+^N$, where λ_+ is the larger eigenvalue of $\mathbf{T}$,

$$\lambda_+ = \mathrm{e}^{\gamma} \left[\cosh\alpha + \sqrt{\sinh^2 \alpha + \mathrm{e}^{-4\gamma}} \right].$$

Finally, substituting the result into Eq. (4.5.16) shows that the mean extension is

$$\langle x \rangle = \frac{Na \sinh\alpha}{\sqrt{\sinh^2 \alpha + \mathrm{e}^{-4\gamma}}}.$$

Note that this reduces to the force-extension relation (4.5.6) of the random walk model in the limit $\gamma \to 0$. The Ising model will be used to model receptor clustering in Sect. 5.3.

Example 4.4 (Persistence Length of a Polymer). In the above models there is a fundamental length scale, namely, the length a of each link, which is called the *Kuhn length*. Roughly speaking, one can view the Kuhn length as the length over which a polymer is essentially straight. In order to consider properties of polymers on length scales smaller than the Kuhn length, which are important for strong deformations, it is more convenient to consider a continuum model of a polymer. One now treats a polymer as a continuous curve in 3D space, parameterized by arc length s (see

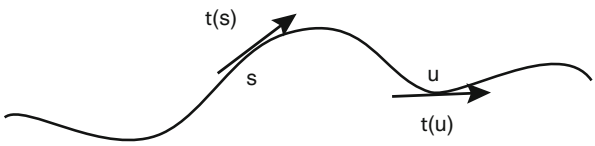


Fig. 4.30: Persistence length of a polymer represented as a *continuous curve*

Fig. 4.30). The Kuhn length a of the freely jointed chain model can then be related to the *persistence length* ξ_p of the continuous model, with the latter defined according to the correlation length over which tangent–tangent correlations decay along the chain:

$$\langle \mathbf{t}(s) \cdot \mathbf{t}(u) \rangle = \mathrm{e}^{-|s-u|/\xi_p}. \tag{4.5.17}$$

For example, the DNA of viruses such as λ-phage has a contour length of 16.6 µm and a persistence length of $\xi_p \approx 50$ nm at room temperature. One can derive a relationship between a and ξ_p by considering the mean square of the end-to-end vector

$$\mathbf{r} = \int_0^L \mathbf{t}(s) ds.$$

One finds that

$$\begin{aligned} \langle \mathbf{r}^2 \rangle &= \left\langle \int_0^L \mathbf{t}(s) ds \cdot \int_0^L \mathbf{t}(u) du \right\rangle = \int_0^L ds \int_0^L du \, \langle \mathbf{t}(s) \cdot \mathbf{t}(u) \rangle \\ &= 2 \int_0^L ds \int_s^L du \, \mathrm{e}^{-(u-s)/\xi_p} \approx 2 \int_0^L ds \int_0^\infty dx \, \mathrm{e}^{-x/\xi_p} = 2L\xi_p, \end{aligned}$$

for $L \gg \xi_p$. Carrying out an analogous calculation for the freely jointed chain, we have

$$\begin{aligned} \langle \mathbf{r}^2 \rangle &= \left\langle \sum_{j=1}^N \mathbf{a}_j \cdot \sum_{k=1}^N \mathbf{a}_k \right\rangle = \sum_{j,k=1}^N \langle \mathbf{a}_j \cdot \mathbf{a}_k \rangle \\ &= Na^2 + \sum_{j \neq k} \langle \mathbf{a}_j \cdot \mathbf{a}_k \rangle = Na^2, \end{aligned}$$

since $\langle \mathbf{a}_j \cdot \mathbf{a}_k \rangle = 0$ for $j \neq k$. Comparison of the two models shows that $a = 2\xi_p$. Note that the continuous model is the starting point for a more detailed analysis of the elastic properties of polymers using the theory of elastic beams or rods [53].

4.6 Exercises

Problem 4.1 (Single-stranded polymer). Consider the kinetic equations for a single-stranded polymer:

$$\frac{dX_n}{dt} = \varepsilon X_{n+1}(t) + \pi X_{n-1}(t) - [\varepsilon + \pi]X_n(t), \quad n > 0$$

and

$$\frac{dX_0}{dt} = \varepsilon X_1(t) - \pi X_0(t),$$

where X_0 is the concentration of monomer and $X_n, n > 0$, is the concentration of filaments of length $n+1$.

(a) Derive the steady-state solution

$$X_n = \left(\frac{\pi}{\varepsilon}\right)^n X_0.$$

Hence, show that the steady-state concentrations satisfy the equilibrium law of mass action for the reversible reaction

$$X_n + X_0 \underset{\varepsilon}{\overset{\pi_0}{\rightleftharpoons}} X_{n+1},$$

with $\pi = \pi_0 X_0$.

(b) Show that the average length of polymers (not counting the monomers) is

$$\langle n \rangle \equiv \frac{\sum_{n=1}^{\infty}(n+1)X_n}{\sum_{n=1}^{\infty} X_n} = 1 + \frac{K}{K-1}, \quad K = \frac{\pi}{\varepsilon}$$

and that the total concentration of subunits is

$$A \equiv \sum_{n=0}^{\infty}(n+1)X_n = \frac{X_0 K^2}{(K-1)^2}.$$

Problem 4.2 (Double-stranded polymer). Consider a double-stranded filament consisting of two protofilaments whose ends can take on two basic types of configuration as shown in Fig. 4.31. Let $P_n(a,0)$ denote the probability that the spatial separation of the protofilament tips is $a+nd$, where d is the length of a monomer, and let $P_n(0,d-a)$ denote the probability that the spatial separation is $nd+d-a$ (cases (a) and (b) of Fig. 4.31). The corresponding configurations are labeled by $(a+nd,0)$ and $(0,d-a+nd)$. The dissociation constants associated with the reversible binding of a monomer will depend on which configuration it binds to and whether or not it binds to the longer or shorter protofilament. Let ΔF_m be the change in free energy due to the removal of a monomer from solution, let $2l\Delta F_{\text{lat}}/d$ be the

lateral binding energy when the added monomer overlaps the other protofilament by an amount l, and denote the binding energy due to attachment of a monomer to the end of a protofilament by ΔF_{long}. There are then four possible reactions involving the binding of an additional monomer, with the following dissociation constants:

$$K_1 \equiv \frac{w_1}{u_1} = c\mathrm{e}^{(2\Delta F_{\text{lat}} a/d + \Delta F_{\text{long}} + \Delta F_m)/k_B T}$$
$$K_2 \equiv \frac{w_2}{u_2} = c\mathrm{e}^{(2\Delta F_{\text{lat}}(1-a/d) + \Delta F_{\text{long}} + \Delta F_m)/k_B T}$$
$$K_3 \equiv \frac{w_3}{u_3} = c\mathrm{e}^{(2\Delta F_{\text{lat}} + \Delta F_{\text{long}} + \Delta F_m)/k_B T}$$
$$K \equiv \frac{w}{u} = c\mathrm{e}^{(\Delta F_{\text{long}} + \Delta F_m)/k_B T},$$

where u_j, w_j are forward and backward reaction rates and c is a fixed concentration of monomers in solution.

(a) Sketch the type of polymer configuration that binds an additional monomer in each of the four reactions, i.e., determine whether the configuration is $(a+nd, 0)$ or $(0, d-a+nd)$ and determine the allowed values of n for each case.
(b) Construct a reaction diagram and show that at steady state

$$P_n(a,0) = \left(\frac{u+w_3}{u_3+w}\right)^n P_0(a,0)$$
$$P_n(0,d-a) = \left(\frac{u+w_3}{u_3+w}\right)^n P_0(0,d-a)$$
$$P_0(0,d-a) = \frac{u_1+w_2}{u_2+w_1} P_0(a,0),$$

which is supplemented by the normalization condition

$$\sum_{n=0}^{\infty} [P_n(a,0) + P_n(0,d-a)] = 1.$$

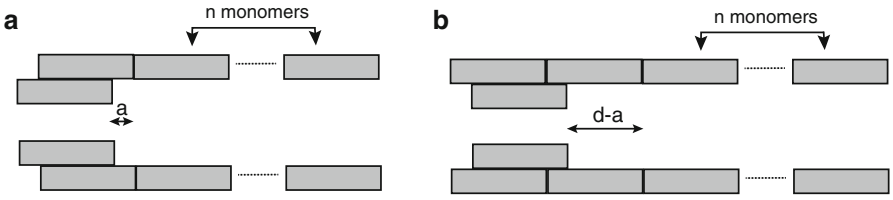


Fig. 4.31: Different configurations of a double-stranded filament. (**a**) One of the protofilaments has a tip at a distance $nd+a$ from the other protofilament with $n \geq 0$. (**b**) One of the protofilaments has a tip at a distance $nd+d-a$ from the other protofilament with $n \geq 0$. The identity of the longer protofilament is not important

(c) Using part (b) and assuming

$$\beta = \frac{u + w_3}{u_3 + w} < 1,$$

sum the geometric series to show that

$$P_0(a) = \frac{1-\beta}{1+\alpha}, \quad P_0(d-a) = \frac{\alpha(1-\beta)}{1+\alpha}, \text{ where } \alpha = \frac{u_1 + w_2}{u_2 + w_1}.$$

(d) The mean growth velocity V is defined as the mean rate of growth of the population of polymers due to polymerization minus the mean rate of shrinkage due to depolymerization. Using this definition, derive the following formula for V:

$$\begin{aligned} V = {} & du + (d-a)u_1 P_0(a) + au_2 P_0(d-a) - dw[1 - P_0(a) - P_0(d-a)] \\ & - (d-a)w_1 P_0(d-a) - aw_2 P_0(a,0). \end{aligned}$$

Hence, show that

$$V = d\left[u - w\beta + (1-\beta)\frac{u_1 u_2 - w_1 w_2}{u_1 + u_2 + w_1 + w_2}\right].$$

Problem 4.3 (Spatial polymerization of a filament). Suppose that a polymer filament is placed in a cylinder with uniform cross section A. Suppose that the monomers within the tube can undergo diffusion along the axis of the tube, which is taken to be the x-axis. Let $x_\pm(t)$ denote the positions of the $\pm$ ends of the filament within the tube. The apparent velocities of these ends due to polymerization/depolymerization are

$$\begin{aligned} \frac{dx_+}{dt} &= v_+ = l[k_{\text{on}}^+ a(x_+,t) - k_{\text{off}}^+] \\ \frac{dx_-}{dt} &= v_- = -l[k_{\text{on}}^- a(x_-,t) - k_{\text{off}}^-]. \end{aligned}$$

The ends of the filament act as sources or sinks for the monomer, so that the monomer concentration $a(x,t)$ along the axis satisfies the inhomogeneous diffusion equation

$$\frac{\partial a}{\partial t} = D\frac{\partial a^2}{\partial x^2} - \gamma[\delta(x-x_+)v_+ - \delta(x-x_-)v_-], \quad \gamma = \frac{1}{Al}.$$

(a) Derive the diffusion equation by considering conservation of monomer passing through an infinitesimal volume $A\Delta x$ centered about either end of the filament. Explain the minus sign in the definition of v_-.

(b) Suppose that the tube is infinitely long and

$$a(x,t) \to \alpha, \quad x \to \pm\infty.$$

Look for a traveling wave solution in which the filament maintains a fixed length L and $v_\pm = v$, where v is the speed of the wave. That is, set $x_+ = vt, v_- = vt - L$ and go to a moving frame $z = x - vt$ with $a(x,t) = \mathscr{A}(z)$ such that

$$-v\frac{d\mathscr{A}}{dz} = D\frac{d^2\mathscr{A}}{dz^2} + v\gamma[\delta(z+L) - \delta(z)].$$

Explicitly solve this equation by matching the solution at the points $z = -L, 0$. In particular, show that

$$\mathscr{A}(-L) = \alpha, \quad \mathscr{A}(0) = \alpha - 1 + \mathrm{e}^{-\gamma vL/D}.$$

(c) Substituting for $\mathscr{A}$ in the expressions for $v_\pm$ and setting $v_+ = v_- = v$, determine v and L. Show that a physical solution only exists if

$$\alpha > \frac{k_{\rm off}^+ + k_{\rm off}^-}{k_{\rm on}^+ + k_{\rm on}^-}.$$

Problem 4.4 (Computer simulations: polymerization).

(a) Use Euler's direct method to solve the ODE corresponding to the master equation (4.1.1) for polymerization at one end:

$$\frac{dP_n}{dt} = \varepsilon P_{n+1}(t) + \pi P_{n-1}(t) - [\varepsilon + \pi]P_n(t), \quad n > 0$$

with $P_n(0) = \delta_{n,10}$, $\varepsilon = 0.7$, $\pi = 0.4$. Take $t = [0,2]$.

(b) Use the Gillespie algorithm (Sect. 6.8) to generate sample paths for the length $N(t)$ of the polymer. The two reactions are $n \to n+1$ at rate π and $n \to n-1$ at rate ε. By averaging over sample paths, compare the histogram of $N(T)$ with the distribution $P_n(T)$ for $T = 2$.

(c) Does the histogram of $N(T)$ appear to converge to a stationary distribution for large T and $\varepsilon = 0.7$, $\pi = 0.4$? What about the case $\varepsilon = 0.4$, $\pi = 0.7$?

Problem 4.5 (Polymerization ratchet). Consider a Brownian particle moving in the ratchet potential

$$\mathscr{F}(x) = Fx - n\Delta G, \quad na < x < (n+1)a.$$

Following the analysis of Sect. 4.2, we obtain the equation

$$\frac{d}{dx}\left(\mathrm{e}^{\mathscr{V}(x)/k_BT}\hat{p}_0(x)\right) = -\frac{\hat{J}_0}{D_0}\mathrm{e}^{\mathscr{V}(x)/k_BT}.$$

for the stationary distribution $\hat{p}_0(x) = \sum_{n=-\infty}^{\infty} p_0(x+na)$.

(a) Integrate the above equation from 0^+ to x, $0 < x < a$, and impose the matching condition

$$\lim_{x\to a^+} \hat{p}_0(x)\mathrm{e}^{\mathscr{F}(x)} = \lim_{x\to a^-} \hat{p}_0(x)\mathrm{e}^{\mathscr{F}(x)}$$

together with periodicity $\hat{p}_0(a^+) = \hat{p}_0(0^+)$. Hence show that

$$\hat{p}_0(x) = \frac{\hat{J}_0 k_B T}{F D_0}\left[\mathscr{A}\mathrm{e}^{-Fx/k_BT} - 1\right],$$

with

$$\mathscr{A} = \frac{\mathrm{e}^{\Delta G/k_BT} - 1}{\mathrm{e}^{(\Delta G - Fa)/k_BT} - 1}.$$

(b) Explain the matching condition used in part (a).

(c) Determine the constant flux $\hat{J}_0$ using the normalization condition $1 = \int_0^a \hat{p}_0(x)dx$. Hence show that the speed of growth $v = \hat{J}_0 a$ is given by

$$v = D_0 \frac{F^2 a}{(k_B T)^2}\left[\mathscr{A}\left(1 - \mathrm{e}^{-Fa/k_BT}\right) - \frac{Fa}{k_BT}\right]^{-1}.$$

(d) Show that in the regime $\Delta G \gg Fa$ and $k_BT \gg Fa$,

$$v \approx 2D_0/a.$$

Problem 4.6 (Translocation ratchet). The FP equation for the translocation ratchet takes the form

$$\frac{\partial p}{\partial t} + \frac{\partial J}{\partial x} = 0, \quad J = -\frac{DF}{k_BT}p - D\frac{\partial p}{\partial x}.$$

Suppose that each ratchet site can exist in two states that are in equilibrium

$$S_0 \underset{k_{\rm off}}{\overset{k_{\rm on}}{\rightleftharpoons}} S_1,$$

with only S_1 ratcheted. The FP equation is then supplemented by the boundary conditions

$$J(0,t) = J(\delta,t), \quad p(\delta) = (1-\pi)p(0), \quad \pi = \frac{k_{\rm on}}{k_{\rm on} + k_{\rm off}}.$$

Show that the velocity of translocation is

$$v = \delta J_0 = \frac{2D}{\delta}\left[\frac{\omega^2/2}{\dfrac{\mathrm{e}^{\omega} - 1}{1 - K(\mathrm{e}^{\omega} - 1)} - \omega}\right].$$

Problem 4.7 (Cooperative model of motor transport). Consider the cooperative model of motor transport in the absence of a load force, for which the cargo attachment and detachment rates are given by

$$\gamma_n = n\gamma_0, \quad \pi_n = (N-n)\pi_0.$$

(a) Show that the steady-state distribution is

$$P_n = P_0 \frac{N!}{(N-n)!n!}\left(\frac{\pi_0}{\gamma_0}\right)^n, \quad P_0 = \left(1+\frac{\pi_0}{\gamma_0}\right)^{-N}.$$

(b) By constructing the generating function $G(s) = \sum_{n=1}^{N} e^{sn} P_n/(1-P_0)$, derive an expression for the mean number of bound motors using $N_b = G'(0)$ and show that for large N,

$$N_b \approx N \frac{\pi_0/\gamma_0}{1+\pi_0/\gamma_0}.$$

Problem 4.8 (Energetics of membrane tethering). Consider a membrane nanotube of length L and radius r, pulled from a spherical vesicle of radius R by an applied force F (see Fig. 4.21). For simplicity, assume that the membrane is in the elastic regime. In order to compute the force F and radius r, we need to consider the energetics associated with bending and stretching the membrane. First, membranes with higher curvature require more bending energy and the rate at which energy changes with curvature is given by the bending stiffness κ. Treating the system as the union of a sphere, cylinder, and hemisphere (at the tip), respectively, the total bending energy is

$$E_{\text{bend}} = 12\pi\kappa + \pi\kappa\frac{L}{r}.$$

The energy associated with changing the area by an amount $\Delta A = A - A_0$, where $A = 4\pi R^2 + 2\pi rL$ and $A_0 = 4\pi R_0^2$, is

$$E_{\text{stretch}} = \frac{K_a}{2}\frac{\Delta A^2}{A_0},$$

where K_a is the elastic rigidity. Another contribution to the total energy arises from the pressure difference Δp between the inside and outside of the vesicle multiplied by the volume:

$$E_{\text{p}} = -\Delta p\left(\frac{4}{3}\pi R^3 + \pi L r^2\right).$$

The total energy of the system is thus

$$E_{\text{tot}} = E_{\text{bend}} + E_{\text{stretch}} + E_{\text{p}} - FL,$$

where FL is the work done by the load. (Note the contributions to the area and volume from the hemispherical tip region have been neglected.)

(a) In order to determine the equilibrium shape of the vesicle + tube, minimize E_{tot} with respect to the three variables r, R, L.
(b) Using the fact that the membrane tension in the elastic regime satisfies

$$\tau = K_a \frac{\Delta A}{A_0},$$

derive the Laplace–Young relation $\Delta p = 2\tau/R$ from the condition $\partial E_{\text{tot}}/\partial R = 0$. This then justifies neglecting the Δp terms in the equations $\partial E_{\text{tot}}/\partial r = 0$ and $\partial E_{\text{tot}}/\partial L = 0$ since $r \ll R$. Hence, derive the equations

$$F = 2\pi\sqrt{2\kappa\tau}, \quad r = \sqrt{\kappa/2\tau}.$$

(c) Using part (b) and the approximation

$$\tau \approx \tau_0 + K_a \frac{rL}{2R_0^2},$$

derive the following force–displacement relation in the elastic regime:

$$\frac{L}{L_e} = \frac{F}{F_0}\left[\left(\frac{F}{F_0}\right)^2 - 1\right], \tag{4.6.18}$$

where $L_e = R_0^2 F_0^3/(8\pi^3\kappa^2 K_a)$ and $F = F_0$ when $\tau = \tau_0$.

Problem 4.9 (Spontaneous oscillations of collective molecular motors). Consider the system of rigidly linked molecular motors shown in Fig. 4.26. Assuming that the system is incommensurate, that is, l/q is an irrational number and the potential V_2 is a constant, Eq. (4.4.35) reduces to the form

$$\frac{\partial P_1}{\partial t} + v\frac{\partial P_1}{\partial \xi} = -(\omega_1 + \omega_2)P_1 + \frac{\omega_2}{l},$$

where $P_1(\xi, t)$ is the probability density that there is a bound molecular motor at ξ, and the force-balance equation (4.4.36) for the displacement $Y(t)$ of the backbone becomes

$$F_{\text{ext}} = \mu v(t) + KY(t) + \int_0^l P_1 \partial_\xi V_1 d\xi, \quad v(t) = \dot{Y}(t).$$

(a) Let

$$P_1(\xi, t) = R(\xi) + Q(\xi, t), \quad R(\xi) = l^{-1}\omega_2/(\omega_1 + \omega_2),$$

and assume a T-periodic solution of the form

$$Q(\xi, t) = \sum_{k\neq 0} Q_k(\xi) e^{ik\omega t}, \quad v(t) = \sum_{k\neq 0} v_k e^{ik\omega t},$$

where $\omega = 2\pi/T$, and set $F_{\text{ext}} = 0$. Show that the coefficients $Q_k(\xi)$ satisfy

$$Q_k(\xi) = -\sum_{l,m} \frac{\delta_{k,l+m}}{\alpha(\xi) + i\omega k} v_l \left[Q'_m(\xi) + \delta_{m,0} R'(\xi) \right], \quad \alpha(\xi) = \omega_1(\xi) + \omega_2(\xi).$$

(b) Suppose that the velocity $v(t) \ll 1$ (relative to typical length and time scales of the system), so that one can perform a perturbation analysis about the steady state, ($P_1(\xi) = R(\xi)$, $Y(t) = Y_0$, $v = 0$), with respect to the Fourier coefficients v_k. Introduce the perturbation expansion

$$Q_k(\xi) = \sum_l Q^{(1)}_{kl}(\xi) v_l + \sum_{l,m} Q^{(2)}_{klm}(\xi) v_l v_m + \dots$$

and substitute into the differential equation for the coefficients $Q_k(\xi)$. Derive the recursion relation

$$Q^{(n)}_{k,k_1,\dots,k_n}(\xi) = -\sum_l \frac{\delta_{k,k_n+l}}{\alpha(\xi) + i\omega k} \partial_\xi Q^{(n-1)}_{l,k_1,\dots,k_{n-1}}(\xi)$$

for $n > 1$ with

$$Q^{(1)}_{kl}(\xi) = -\frac{\delta_{k,l}}{\alpha(\xi) + i\omega k} R'(\xi).$$

(c) Show that the force-balance equation with $F_{\text{ext}} = 0$ can be Taylor expanded as

$$0 = \sum_l F^{(1)}_{kl} v_l + \sum_{lm} F^{(2)}_{klm} v_l v_m + \dots,$$

with

$$F^{(1)}_{kl} = \delta_{k,l} \left(\mu + \frac{K}{i\omega k} - \int_0^l \frac{R'(\xi) V'_1((\xi)}{\alpha(\xi) + i\omega k} d\xi \right),$$

and for $n > 1$

$$F^{(n)}_{k,k_1,\dots,k_n} = \int_0^l Q^{(n)}_{k,k_1,\dots,k_n}(\xi) V'_1(\xi) d\xi.$$

Note that $F^{(1)}_{11}(\omega) = 0$ recovers the instability condition of the steady state, which occurs at a critical frequency $\omega = \omega_c$.

(d) Use parts (b) and (c) to prove that $F_{k,k_1,\dots,k_n} = 0$ unless $k = k_1 + \dots + k_n$. Close to the bifurcation point ($\omega = \omega_c, v = 0$), the dominant terms in the Fourier expansion are $v_{\pm 1}$ and $v_{\pm 2}$. Dropping higher-order terms $v_{\pm n}, n > 2$, use the force-balance equation to show that $v_{\pm 1}, v_{\pm 2}$ satisfy the pair of equations

$$0 = F^{(1)}_{11} v_1 + G^{(2)} v_{-1} v_2 + G^{(3)} v_1^2 v_{-1}, \quad 0 = F^{(1)}_{22} v_2 + F^{(2)}_{211} v_1^2,$$

where $v_{-1} = \overline{v}_1$ and

$$G^{(2)} = F^{(2)}_{1,2,-1} + F^{(2)}_{1,-1,2}, \quad G^{(3)} = F^{(3)}_{1,1,1,-1} + F^{(3)}_{1,1,-1,1} + F^{(3)}_{1,-1,1,1}.$$

Deduce that the amplitude of the spontaneous oscillation satisfies

$$|v_1|^2 = -F_{11}^{(1)}/\tilde{G}^{(3)}, \quad \tilde{G}^{(3)} = G^{(3)} - F_{211}^{(2)} G^{(2)}/F_{22}^{(1)}.$$

Since $|v_1|^2$ is real, setting the imaginary part of $F_{11}^{(1)}/\tilde{G}^{(3)}$ to zero determines ω.

Problem 4.10 (Random walk model of a 1D polymer). Consider a 1D random polymer consisting of N links, each of which is described by a vector $\mathbf{a}_j = \pm a\mathbf{e}_x$ with equal probability $p_\pm = 1/2$, where $\mathbf{e}_x$ is the unit vector in the x-direction (see Fig. 4.28). One end of the polymer is fixed at $x = 0$. Formulate this model as a random walk problem, with the number of segments N analogous to the number of time steps and the end-to-end displacement analogous to position on a 1D lattice after the Nth step. Hence, show that the probability distribution of R, where $\sum_{j=1}^{N} \mathbf{a}_j = R\mathbf{e}_x$ is approximately given by

$$P_N(R) \sim \frac{1}{\sqrt{2\pi N}} \mathrm{e}^{-R^2/2Na^2}.$$

Chapter 5
Sensing the Environment: Adaptation and Amplification in Cells

One important requirement of sensory eukaryotic cells and single-cell organisms such as bacteria is detecting weak signals in noisy extracellular environments. As we briefly discussed in Sect. 2.4.2 within the context of bacterial chemoreception, there are fundamental limits to the strength of signal that can be detected. However, even if a weak signal is detected, it is necessary for some form of *amplification* to occur in order that the signal is not lost in subsequent stages of processing within the cell. Moreover, it is advantageous for a cell to be able to shift its response so that it always operates in a regime of maximal gain, that is, it is able to respond to small changes in signal irrespective of the mean strength of the signal—a process known as *adaptation*. In this chapter we explore these issues in more detail. We begin by considering the Berg–Purcell limit of biochemical signaling [40] and a subsequent modification of the original result that applies outside the diffusion-limited regime (Sect. 5.1). In Sect. 5.2 we review an alternative approach to estimating the precision of biochemical sensors that is based on the fluctuation–dissipation theorem of statistical mechanics. In Sect. 5.3 we return to the problem of bacterial chemotaxis, which is a canonical system used to explore the sensitivity of biochemical sensors to environmental signals. We describe some of the biochemical signaling networks responsible for amplification via receptor clustering (cooperativity) and for adaptation. We also analyze some simple PDE models of bacterial chemotaxis. Finally, in Sect. 5.4, we consider how amplification and adaptation occur in hair cells of the inner ear via active mechanotransduction. Here the interactions between the mechanical properties of transduction elements, the action of myosin motors, and Ca^{2+} signaling allow a hair cell to operate close to a Hopf bifurcation point for the onset of spontaneous oscillations. This, in turn, provides the basis for active signal processing such as amplification and frequency tuning.

P.C. Bressloff, *Stochastic Processes in Cell Biology*, Interdisciplinary Applied Mathematics 41, DOI 10.1007/978-3-319-08488-6_5

5.1 Physical Limits of Biochemical Signaling due to Noise

We begin by presenting the original derivation of the Berg–Purcell limit for the precision with which the external concentration c of some ligand such as a chemoattractant can be inferred from the time-averaged occupancy of a single receptor embedded in the cell membrane [37, 40]. Let us denote the time-dependent state of the receptor by $n(t)$ with $n(t) = 1$ ($n(t) = 0$) if the receptor's binding site is occupied (unoccupied) by a single ligand molecule. In thermodynamic equilibrium at some given concentration c, the time-averaged occupation $\bar{n}$ is

$$\bar{n} = \frac{c}{c + K_d}, \tag{5.1.1}$$

where K_d is the dissociation constant for ligand binding/unbinding (see Sect. 3.1). After a molecule is bound to the receptor, the probability of detachment in an interval dt is dt/τ_b where τ_b is the inverse of the unbinding rate. Suppose that the binding site is treated as a circular disk of radius a. From the analysis of diffusion-limited reactions in Sect. 2.4, we know that the diffusive flux into the dic is $4Dac$, where D is the ligand diffusivity. At equilibrium the rate at which a molecule binds to the receptor must balance the rate of escape:

$$\frac{\bar{n}}{\tau_b} = 4(1 - \bar{n})Dac. \tag{5.1.2}$$

In particular, since $\bar{n} = 1/2$ when $c = K_d$, it follows that

$$\tau_b = \frac{1}{4DaK_d}. \tag{5.1.3}$$

The information about the surrounding concentration c available to the cell is the function $n(t)$ sampled over a time interval τ_{avg}. Let

$$n_* = \frac{1}{\tau_{\text{avg}}} \int_{t_0}^{t_0+\tau_{\text{avg}}} n(t)dt \tag{5.1.4}$$

be the cell's estimate of $\bar{n}$, with t_0 the time when sampling begins. It follows from equation (5.1.1) that the corresponding estimate of c is

$$c_* = \frac{n_*}{1 - n_*} K_d. \tag{5.1.5}$$

We would like to determine the error in such an estimate. From the definition of n_*, we have

$$n_*^2 = \frac{1}{\tau_{\text{avg}}^2} \int_{t_0}^{t_0+\tau_{\text{avg}}} \int_{t_0}^{t_0+\tau_{\text{avg}}} n(t)n(t')dtdt',$$

which implies that

$$\langle n_*^2 \rangle = \frac{1}{\tau_{\text{avg}}^2} \int_0^{\tau_{\text{avg}}} \int_0^{\tau_{\text{avg}}} C(t-t')dt\,dt', \tag{5.1.6}$$

where $C(\tau)$ is the correlation function

$$C(\tau) = \langle n(t)n(t+\tau) \rangle. \tag{5.1.7}$$

It is assumed that the stochastic process is stationary so that C only depends on the time difference and the expression for $\langle n_*^2 \rangle$ is independent of the initial time t_0. The correlation function satisfies the differential equation

$$\tau_b \frac{dC}{d\tau} = -C(\tau) + (\bar{n} - C(\tau)) \frac{\bar{p}}{1-\bar{p}}. \tag{5.1.8}$$

In order to derive this equation, one imagines making a large number N of measurements $n(t_0)n(t_0+\tau)$ over a random set of initial sampling times t_0. Since $n(t_0)n(t_0+\tau) = 1$ only if $n(t_0) = n(t_0+\tau) = 1$, otherwise $n(t_0)n(t_0+\tau) = 0$, one can keep track of changes in the number of nonzero pair-wise measurements when $\tau \to \tau + d\tau$, which leads to equation (5.1.8) [40]. Integrating this equation and requiring that $C(0) = \bar{n}$ gives

$$C(\tau) = \bar{n}^2 + \bar{n}(1-\bar{n})\mathrm{e}^{-|\tau|/(1-\bar{n})\tau_b}. \tag{5.1.9}$$

Substituting for $C(\tau)$ into equation (5.1.6) yields the result

$$\begin{aligned} \langle n_*^2 \rangle =& \frac{1}{\tau_{\text{avg}}^2} \int_0^{\tau_{\text{avg}}} \int_0^{\tau_{\text{avg}}} \left[\bar{n}^2 + \bar{n}(1-\bar{n})\mathrm{e}^{-|t-t'|/(1-\bar{n})\tau_b} \right] dt\,dt' \\ =& \bar{n}^2 + \frac{2}{\tau_{\text{avg}}} \bar{n}(1-\bar{n})^2 \tau_b. \end{aligned}$$

Given the unbiased estimate $\langle n_* \rangle = \bar{n}$, we see that the RMS error δn in our estimate of $\bar{n}$ satisfies

$$\frac{\delta n}{\bar{n}} = \frac{1}{n_*} \sqrt{\langle n_*^2 \rangle - \langle n_* \rangle^2} = \sqrt{\frac{2}{\bar{n}\tau_{\text{avg}}} (1-\bar{n})^2 \tau_b},$$

which, combined with equation (5.1.2), reduces to

$$\frac{\delta n}{\bar{n}} = \sqrt{\frac{1-\bar{n}}{2Dac\tau_{\text{avg}}}}. \tag{5.1.10}$$

If we now use the small noise approximation

$$\frac{\delta c}{c} = \frac{dc}{d\bar{n}} \frac{\delta n}{c} = \frac{K_d}{(1-\bar{n})^2} \frac{1-\bar{n}}{K_d \bar{n}} \delta n = \frac{1}{1-\bar{n}} \frac{\delta n}{\bar{n}},$$

we deduce that

$$\frac{\delta c}{c} = \sqrt{\frac{2}{4Dac(1-\bar{n})\tau_{\rm avg}}}. \tag{5.1.11}$$

Equations (5.1.10) and (5.1.11) express the fundamental limits of Berg and Purcell [40]. If the binding site is modeled as a sphere rather than a disk, then there is a different geometric factor so $4Dac \to 2\pi aDc$.

One of the simplifying assumptions of the above derivation is that reactions are *diffusion-limited*. That is, as soon as a ligand molecule comes into contact with the receptor, it is absorbed. In other words, the rates of binding and unbinding $k_\pm \to \infty$ with $k_-/k_+ = K_d$. Berg and Purcell argued that their result also holds for reactions that are not diffusion-limited (finite $k_\pm$) since, if a ligand molecule fails to bind, then it will rapidly keep re-colliding with the receptor until it does eventually bind. Such a process can be captured by rescaling the radius a. However, this ignores the possibility that after unsuccessfully binding, the specific ligand molecule diffuses back into the bulk, and a different ligand molecule subsequently binds. Similarly, a ligand molecule that has just dissociated from the receptor could either rapidly rebind or diffuse away into the bulk. Recently, a detailed study of the Berg–Purcell problem outside the diffusion-limited regime has been carried out by Kaizu et al. [314], who show that there are now two contributions to the RMS error:

$$\frac{\delta c}{c} = \sqrt{\frac{1}{2\pi Dac(1-\bar{n})\tau_{\rm avg}} + \frac{2}{k_+ c(1-\bar{n})\tau_{\rm avg}}}. \tag{5.1.12}$$

The first term recovers the fundamental limit of Berg and Purcell applied to a sphere rather than a circular disk, whereas the second term takes into account the variability that results from the receptor–ligand binding kinetics; the latter vanishes in the limit $k_+ \to \infty$. Interestingly, a similar result has been obtained by Bialek and Setayeshgar [45] using a very different approach, which is based on the fluctuation–dissipation (FD) theorem of statistical mechanics:

$$\frac{\delta c}{c} = \sqrt{\frac{1}{\pi Dac\tau_{\rm avg}} + \frac{2}{k_+ c(1-\bar{n})\tau_{\rm avg}}}. \tag{5.1.13}$$

The contribution to uncertainty from binding kinetics agrees with the Kaizu et al. result [314], but the contribution from diffusion differs from both the latter and Berg–Purcell. One possible explanation for this discrepancy is that the FD theorem only takes into account linear correlations. On the other hand, the FD theorem provides a relatively simple, intuitive method for addressing the physical limits of biochemical signaling and will be discussed at length in the next section.

5.2 Fluctuation–Dissipation Theorem

The fluctuation–dissipation (FD) theorem is based on the assumption that the response of a system in thermodynamic equilibrium to a small applied force is the same as its response to a spontaneous fluctuation. Suppose that $x(t)$ represents the linear response of the system to a small external input $h(t)$, such that

$$x(t) = \int_{-\infty}^{\infty} G(\tau)h(t-\tau)d\tau, \quad G(\tau) = 0 \text{ for } \tau < 0, \tag{5.2.1}$$

where $G(t)$ is the linear response function. In a mechanical system $x(t)$ would represent a physical displacement and h an applied force, whereas in a magnetic system $x(t)$ would represent magnetization and h an applied magnetic field. In the latter case, $G(t)$ is known as the magnetic susceptibility. Let $X(t)$ represent the corresponding response of the system to thermal fluctuations. The FD theorem states that the power spectrum $S_X(\omega)$ (defined in Sect. 2.2.5) is related to the Fourier transform of the linear response function $G(t)$ according to

$$S_X(\omega) = \frac{2k_BT}{\omega}\text{Im}[\tilde{G}(\omega)]. \tag{5.2.2}$$

(The FD theorem can be derived from first principles using statistical physics and the observation that at equilibrium $\langle X\rangle = \partial E/\partial h$, where E is the free energy of the system; see Box 5A.) For example, applying the FD theorem to the OU process (Sect. 2.2.3) with $\tilde{G}(\omega) = [\kappa - i\omega\gamma]^{-1}$, we find that

$$S_X(\omega) = \frac{2k_BT\gamma}{\kappa^2 + (\omega\gamma)^2}. \tag{5.2.3}$$

Comparison with the calculated expression for $S_X(\omega)$ in equation (2.2.33) recovers the Einstein relation $D = k_BT\gamma$.

Box 5A. Derivation of the fluctuation–dissipation theorem.

We sketch a proof of the FD theorem for a system close to thermodynamic equilibrium using the Boltzmann–Gibbs distribution introduced in Sect. 1.4 (see also [102]). Imagine that a constant external force h is applied to the system over the time interval $(-\infty, 0]$ and then suddenly switched off at time $t = 0$. Let E_j be the energy of a microstate in the absence of the force and $E_j - hA_j$ be the energy of the state for $h \neq 0$, where A_j is some function of state. Since the system has had an infinite time to relax to equilibrium prior to $t = 0$, we know that before the force is switched off the mean $\langle A(0)\rangle$ is given by

$$\langle A(0)\rangle = \frac{\sum_j A_j \mathrm{e}^{-\beta(E_j - hA_j)}}{\sum_j \mathrm{e}^{-\beta(E_j - hA_j)}}, \quad \beta = \frac{1}{k_B T}.$$

Let $A(t|j)$ be the value of A at time t, which evolves in the absence of the applied force from one of the initial microstates $A_j(0)$ at time $t = 0$. Averaging over these initial states and using the fact that h is infinitesimal, we have

$$\begin{aligned}
\overline{A}(t) = \langle A(t|j)\rangle &= \frac{\sum_j A(t|j)\mathrm{e}^{-\beta(E_j - hA_j)}}{\sum_j \mathrm{e}^{-\beta(E_j - hA_j)}} \\
&\approx \frac{\sum_j A(t|j) e^{-\beta E_j}(1 + \beta h A_j(0) + \ldots)}{\sum_j \mathrm{e}^{-\beta E_j}(1 + \beta h A_j(0) + \ldots)} \\
&\approx \frac{\sum_j e^{-\beta E_j} A(t|j)\left(1 + \beta h A_j(0) - \beta h \frac{\sum_k \mathrm{e}^{-\beta E_k} A_k(0)}{\sum_k \mathrm{e}^{-\beta E_k}}\right)}{\sum_j \mathrm{e}^{-\beta E_j}} \\
&= \langle A(t)\rangle + \beta h\left[\langle A(t)A(0)\rangle - \langle A(t)\rangle\langle A(0)\rangle\right].
\end{aligned}$$

In other words, the relaxation to equilibrium can be related to the autocorrelation function $C(t)$ of spontaneous fluctuations $\delta A(t)$:

$$\overline{A}(t) - \langle A\rangle = \langle \delta A(t)\delta A(0)\rangle = C(t).$$

Let $G(t)$ be the linear response function. It follows that we can also represent the relaxation to equilibrium according to

$$\overline{A}(t) - \langle A\rangle = \int_{-\infty}^{\infty} G(\tau)h(t-\tau)d\tau.$$

Using the particular piecewise constant form for h,

$$\overline{A}(t) - \langle A\rangle = h\int_t^{\infty} G(\tau)d\tau,$$

which implies that

$$G(t) = -\beta\frac{dC(t)}{dt}H(t).$$

Fourier transforming this equation gives

$$\tilde{G}(\omega) = -\beta\int_0^{\infty}\dot{C}(t)\mathrm{e}^{i\omega t}dt = -\beta + i\omega\beta\int_0^{\infty} C(t)\mathrm{e}^{i\omega t}dt.$$

Taking the imaginary part of this equation,

$$\begin{aligned}\mathrm{Im}\,\tilde{G}(\omega) &= \omega\beta \mathrm{Re}\int_0^\infty C(t)\mathrm{e}^{i\omega t}dt = \frac{\omega\beta}{2}\left[\int_0^\infty C(t)\mathrm{e}^{i\omega t} + \int_0^\infty C(t)\mathrm{e}^{-i\omega t}\right]dt \\ &= \frac{\omega\beta}{2}\left[\int_0^\infty C(t)\mathrm{e}^{i\omega t} + \int_0^\infty C(-t)\mathrm{e}^{-i\omega t}\right]dt = \frac{\omega\beta}{2}\int_{-\infty}^\infty C(t)\mathrm{e}^{i\omega t}dt \\ &= \frac{\omega\beta}{2}S(\omega),\end{aligned}$$

where $S(\omega)$ is the power spectrum.

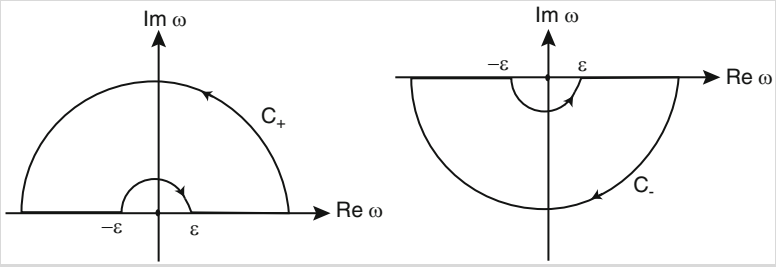


Fig. 5.1: Contours in the complex frequency plane ω

Another version of the FD theorem. This can be obtained by comparing two formulae for the total power. First, integrating equation (2.2.29) with respect to ω and ω' shows that

$$\langle X(0)^2\rangle \equiv \int_{-\infty}^\infty\int_{-\infty}^\infty \langle\tilde{X}(\omega)\tilde{X}(\omega')\rangle\frac{d\omega}{2\pi}\frac{d\omega'}{2\pi} = \int_{-\infty}^\infty S_X(\omega)\frac{d\omega}{2\pi}.$$

On the other hand, integrating (5.2.2) with respect to ω gives

$$\int_{-\infty}^\infty S_X(\omega)\frac{d\omega}{2\pi} = 2k_BT\int_{-\infty}^\infty \frac{\mathrm{Im}[\tilde{G}(\omega)]}{\omega}\frac{d\omega}{2\pi}.$$

Recall that $G(\tau)$ is a causal Green's function, which means that $G(\tau) = 0$ for $\tau < 0$. It follows from $G(\tau) = \int_{-\infty}^\infty \tilde{G}(\omega)\mathrm{e}^{-i\omega\tau}d\omega/2\pi$ that $\tilde{G}(\omega)$ is analytic in the upper-half complex ω-plane. This also means that $\tilde{G}^*(\omega) = \tilde{G}(-\omega)$ is analytic in the lower-half complex plane. Hence, we can rewrite the integral of $\mathrm{Im}[\tilde{G}(\omega)]$ as a sum of contour integrals

$$\int_{-\infty}^\infty \frac{\mathrm{Im}[\tilde{G}(\omega)]}{\omega}\frac{d\omega}{2\pi} = \frac{1}{2i}\oint_{\mathscr{C}_+}\frac{[\tilde{G}(\omega)}{\omega}\frac{d\omega}{2\pi} - \frac{1}{2i}\oint_{\mathscr{C}_-}\frac{\tilde{G}(\omega)}{\omega}\frac{d\omega}{2\pi},$$

where $\mathscr{C}_+$ ($\mathscr{C}_-$) is closed in the upper-half (lower-half) complex plane (see Fig. 5.1). From the calculus of residues, we then have

$$\int_{-\infty}^{\infty} \frac{\mathrm{Im}[\tilde{G}(\omega)]}{\omega} \frac{d\omega}{2\pi} = \frac{1}{2}\tilde{G}(0).$$

We thus obtain another version of the FD theorem

$$\langle X(0)^2 \rangle = k_B T \tilde{G}(0). \tag{5.2.4}$$

In the case of the OU equation, $\tilde{G}(0) = \kappa^{-1}$ and

$$\frac{\kappa}{2}\langle X(0)^2 \rangle = \frac{1}{2}k_B T,$$

which is an expression of the equipartition theorem of statistical mechanics: in equilibrium each degree of freedom has a mean energy of $k_B T/2$.

The FD theorem has traditionally been applied to mechanical and other physical systems within the context of statistical physics. More recently, however, Bialek and colleagues have used the FD theorem to explore a number of important issues in biochemical signaling, including the identification of fundamental physical limits on the sensitivity of biochemical sensors such as receptors [45, 46, 630]. Following [45], consider the simple problem of signaling molecules binding to a single site of a receptor. Ignoring fluctuations, the fractional occupancy of the site, $n(t)$, evolves according to the first-order kinetic equation

$$\frac{dn}{dt} = k_+ c[1 - n(t)] - k_- n(t), \tag{5.2.5}$$

where c is the background concentration of the ligand. The equilibrium law of mass action requires that the rate constants are related through detailed balance (see Sects. 1.4 and 3.1),

$$\frac{k_+ c}{k_-} = \mathrm{e}^{F/k_B T},$$

where F is the free energy associated with binding. Now suppose that thermal noise induces small fluctuations in the binding energy, δF, and in the associated rate constants, $\delta k_\pm$. Taking logs of the detailed balance equation shows that the fluctuations are related according to

$$\frac{\delta k_+}{k_+} - \frac{\delta k_-}{k_-} = \frac{\delta F}{k_B T}.$$

Linearizing (5.2.5) about the equilibrium solution

$$\bar{n} = \frac{k_+ c}{k_+ c + k_-}, \tag{5.2.6}$$

gives

$$\frac{d\delta n}{dt} = -(k_+c + k_-)\delta n + c(1-\bar{n})\delta k_+ - \bar{n}\delta k_-.$$

Combining this with the detailed balance constraint yields the linear equation

$$\gamma\frac{d\delta n}{dt} = -\kappa\delta n + \delta F, \tag{5.2.7}$$

with

$$\gamma = \frac{k_BT}{k_+c(1-\bar{n})}, \quad \kappa = (k_+c + k_-)\gamma = \frac{k_BT}{\bar{n}(1-\bar{n})}. \tag{5.2.8}$$

This has an identical structure to the OU equation but now describes the change in fractional occupation in response to fluctuations in binding free energy rather than the change in position of a particle in response to fluctuating forces in solution. There is an effective damping coefficient γ and an effective spring constant κ. Applying the FD theorem to this system using equations (5.2.3) and (5.2.8) immediately gives

$$S_n(\omega) = \frac{2k_+c(1-\bar{n})}{\omega^2 + (k_+c + k_-)^2} = \langle(\delta n)^2\rangle\frac{2\tau_c}{1+(\omega\tau_c)^2},$$

where, from equation (5.2.4),

$$\langle(\delta n)^2\rangle = \frac{k_BT}{\kappa} = \bar{n}(1-\bar{n}),$$

and $\tau_c = (k_+c + k_-)^{-1}$. The power spectrum is said to have a Lorentzian form and is equivalent to an exponential decay of correlations:

$$\langle\delta n(t)\delta n(t')\rangle = \int_{-\infty}^{\infty} \mathrm{e}^{-i\omega(t-t')}S_n(\omega)\frac{d\omega}{2\pi} = \langle\delta n^2\rangle\mathrm{e}^{-|t-t'|/\tau_c}. \tag{5.2.9}$$

This is easy to check by noting that

$$\begin{aligned}\int_{-\infty}^{\infty}\mathrm{e}^{i\omega t}\mathrm{e}^{-|t|/\tau_c}dt &= \int_0^{\infty}\mathrm{e}^{i\omega t}\mathrm{e}^{-t/\tau_c}dt + \int_{-\infty}^{0}\mathrm{e}^{i\omega t}\mathrm{e}^{t/\tau_c}dt\\ &= \int_0^{\infty}\mathrm{e}^{i\omega t}\mathrm{e}^{-t/\tau_c}dt + \int_0^{\infty}\mathrm{e}^{-i\omega t}\mathrm{e}^{-t/\tau_c}dt\\ &= 2\mathrm{Re}\int_0^{\infty}\mathrm{e}^{i\omega t}\mathrm{e}^{-t/\tau_c}dt = 2\mathrm{Re}\frac{1}{\tau_c^{-1}+i\omega} = \frac{2\tau_c}{1+\omega^2\tau_c^2}.\end{aligned}$$

In summary, the above application of the FD theorem recovers results that could also be obtained using a linear noise approximation of the underlying master equation along analogous lines to the two-state ion channel model of Sect. 3.2. The advantage of the FD theorem is that it makes no assumptions about the underlying microscopic theory (e.g., master equation), other than that the system operates close to thermodynamic equilibrium. It also provides a powerful framework for coupling

chemical reactions with diffusion. In particular, Bialek and Setayeshgar [45] derived an extension of the Berg–Purcell result presented in Sect. 5.1 by coupling the receptor model with diffusion of ligand. Taking the receptor to be at position $\mathbf{x}_0$ and setting $c = c(\mathbf{x},t)$, we have

$$\frac{dn}{dt} = k_+ c(\mathbf{x}_0,t)[1-n(t)] - k_- n(t), \tag{5.2.10}$$

and

$$\frac{\partial c(\mathbf{x},t)}{\partial t} = D\nabla^2 c(\mathbf{x},t) - \delta(\mathbf{x}-\mathbf{x}_0)\frac{dn(t)}{dt}. \tag{5.2.11}$$

The last term on the right-hand side of (5.2.11) takes into account the transfer of a single ligand molecule when it binds to the receptor. The next step is to linearize the equations about the steady-state solution $(\bar{n},\bar{c})$, where $\bar{c}$ is the uniform background concentration. Using a combination of temporal and spatial Fourier transforms, it can be shown that (see Ex. 5.1)

$$\frac{\delta n(\omega)}{\delta F(\omega)} = \frac{k_+\bar{c}(1-\bar{n})}{k_B T}\frac{1}{-i\omega[1+\Sigma(\omega)]+k_+\bar{c}+k_-}, \tag{5.2.12}$$

where

$$\Sigma(\omega) = k_+(1-\bar{n})\int \frac{1}{Dk^2 - i\omega}\frac{d^3k}{(2\pi)^3}. \tag{5.2.13}$$

In bacterial chemotaxis, the receptor occupancy is averaged over a sufficiently long time interval so that high-frequency components of the spectrum are eliminated, and we can thus make the approximation $\Sigma(\omega) \to \Sigma(0)$, with

$$\begin{aligned}\Sigma(0) &= k_+(1-\bar{n})\int \frac{1}{Dk^2}\frac{d^3k}{(2\pi)^3} \\ &= 4\pi k_+(1-\bar{n})\int_0^{\pi/a}\frac{k^2}{Dk^2}\frac{dk}{(2\pi)^3} \\ &= \frac{k_+(1-\bar{n})}{2\pi Da}.\end{aligned}$$

In evaluating the integral we have used spherical polar coordinates and taken the upper bound of the wavenumber $k = \sqrt{\mathbf{k}\cdot\mathbf{k}}$ to be $k = \pi/a$, where a is the size of the receptor. We then have an effective damping coefficient γ and an effective spring constant κ given by

$$\gamma = [1+\Sigma(0)]\gamma_0, \quad \kappa = (k_+c+k_-)\gamma_0, \quad \gamma_0 = \frac{k_B T}{k_+c(1-\bar{n})}. \tag{5.2.14}$$

An application of the FD theorem together with equations (5.2.3) and (5.2.14) shows that

$$S_n(\omega) = \frac{2k_+c(1-\bar{n})[1+\Sigma(0)]}{(\omega[1+\Sigma(0)])^2+(k_+c+k_-)^2} = \bar{n}(1-\bar{n})\frac{2\tau_c}{1+(\omega\tau_c)^2},$$

where

$$\tau_c = \frac{1+\Sigma(0)}{k_+\bar{c}+k_-} = \frac{1-\bar{n}}{k_-} + \frac{\bar{n}(1-\bar{n})}{2\pi Da\bar{c}}. \tag{5.2.15}$$

We have used equation (5.2.6) and the formula for $\Sigma(0)$. The RMS error is then given by

$$\frac{\delta n}{\bar{n}} = \frac{1}{\bar{n}}\sqrt{\frac{S_n(0)}{\tau_{\rm avg}}} = \sqrt{\frac{2(1-\bar{n})^2}{k_-\bar{n}\tau_{\rm avg}} + \frac{(1-\bar{n})^2}{\pi Da\bar{c}\tau_{\rm avg}}}.$$

Using the identity $(1-\bar{n})/k_-\bar{n} = 1/ck_+$ and $\delta c = c\delta n/[\bar{n}(1-\bar{n})]$, we finally obtain equation (5.1.13).

5.3 Bacterial Chemotaxis

We briefly considered the chemotaxis of *E. coli* in Sect. 2.4.2, within the context of diffusion-limited reactions and sensitivity to chemical gradients. Here we consider the role of biochemical signaling networks in bacterial chemotaxis. *E. coli* is one of the most studied organisms in systems biology, exhibiting a number of important

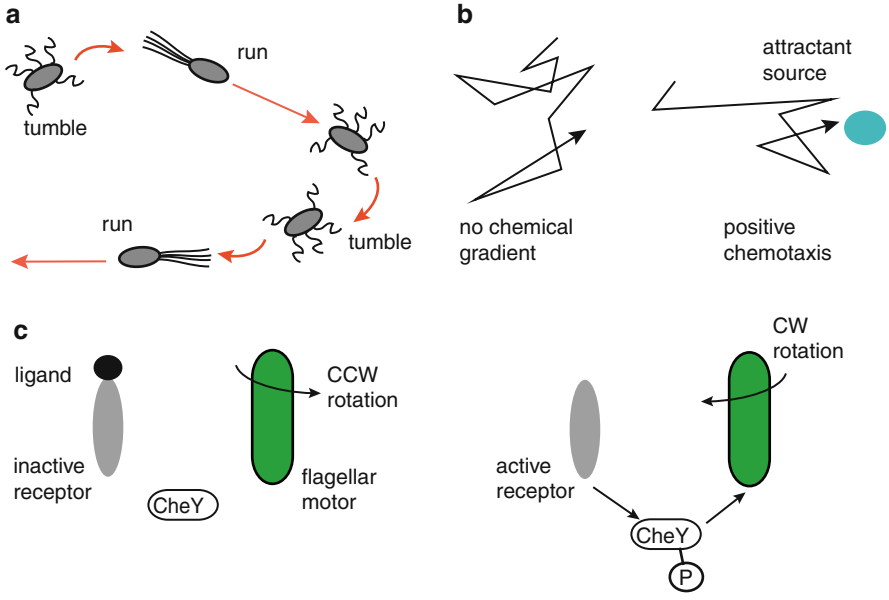


Fig. 5.2: Bacterial chemotaxis. (**a**) A schematic showing the motion of a bacterium that consists of a series of runs and tumbles. (**b**) The sequence of runs and tumbles can be altered by an external chemical gradient so that the motion is biased towards (away from) an attractant (a repellant). (**c**) The switching of a flagellar motor from counterclockwise to clockwise rotation, resulting in a switch from running to tumbling, is controlled by a signaling pathway in which unbinding of a ligand (attractant molecule) from a chemoreceptor in the cell membrane leads to the phosphorylation of CheY, which subsequently binds to the motor and induces the switch

signaling mechanisms including signal amplification, adaptation, and robustness to noise. For an excellent discussion of some of the issues covered here see Chap. 4 of Bialek [44] and the reviews [628, 645]. Many bacteria, including *E. coli*, possess flagella, which are helical polymer filaments that are turned by molecular motors embedded in the cell's membrane. (The axial-asymmetric helical structure of flagella provides a mechanism for swimming at low Reynolds number; see Box 5B.) When all of the flagellar motors are rotating CCW, the helical filaments bundle together and efficiently drive the bacterium in a straight line comprising a single run. On the other hand, if the motors reverse direction, the flagellar bundle flies apart and the bacterium rotates in a random fashion called a tumble. This is illustrated in Fig. 5.2a. Over longer time scales the motion of the bacterium looks like a sequence of straight line trajectories arranged at random angles to each other (see Fig. 5.2b). Tuning of the swimming behavior by environmental signaling molecules allows the bacterium to swim either towards a food source (chemoattractant) or away from a noxious toxin (chemorepellant). These signaling molecules bind to chemoreceptors in the cell membrane that induce dephosphorylation of a downstream signaling molecule CheY which tends to switch the flagellar motors from clockwise to CCW rotation (see Fig. 5.2c).

Box 5B. Swimming at low Reynolds number.

Reynolds number. The flagellar-based swimming mechanism of *E. coli* is one strategy for moving in a fluid at low Reynolds number. In order to understand what this means, it is necessary to consider a little fluid mechanics [38, 122, 517]. Consider a flat body such as a spoon moving in a fluid such as air or water. Roughly speaking, the force necessary to keep the object moving at a constant speed v is $F \sim \mu A dv/dy$, where A is the surface area of the spoon and $v(y)$ is the velocity of different cross sections of fluid at a perpendicular distance y from the object. The constant μ is known as the viscosity of the fluid. (The linear relationship between F and the velocity gradient is characteristic of a Newtonian fluid such as air or water.) Suppose that l is a characteristic size of the object. Using dimensional analysis, the viscous force $\mu A dv/dy$ will scale as μlv, whereas the inertial force mdv/dt due to the fluid's momentum will scale as $\rho l^2 v^2$ where ρ is the density of the fluid. The ratio of these two forces is characterized by a single dimensionless parameter known as the Reynolds number (Re):

$$\mathrm{Re} = \frac{\rho l^2 v^2}{\mu l v} = \frac{\rho l v}{\mu}. \tag{5.3.1}$$

When $\mathrm{Re} \gg 1$ inertial forces dominate, whereas viscous forces dominate when $\mathrm{Re} \ll 1$. Using typical length and velocity scales for humans and bacteria swimming in water, one finds that $\mathrm{Re} \sim 10^4$ for humans and

Re $\sim 10^{-3} - 10^{-5}$ for bacteria. Note that the Reynolds number can also be obtained from the *Navier–Stokes equation*, which is the governing equation in fluid dynamics and is associated with conservation of momentum. In the case of an incompressible fluid ($\nabla \cdot \mathbf{v} = 0$), the Navier–Stokes equation is

$$-\nabla p + \mu \nabla^2 \mathbf{v} = \rho \frac{\partial \mathbf{v}}{\partial t} + \rho(\mathbf{v} \cdot \nabla)\mathbf{v}.$$

The terms on the left-hand side represent pressure and viscous terms, and the terms on the right-hand side correspond to inertial terms. After non-dimensionalizing the Navier–Stokes equation, the right-hand side is multiplied by Re, so that for Re$\ll 1$, the Navier–Stokes equation reduces to the time-independent equation

$$\mu \nabla^2 \mathbf{v} = \nabla p. \tag{5.3.2}$$

Scallop theorem. One of the immediate consequences of swimming at low Reynolds number is that the net forces acting on a body must at all times be zero, since they cannot be counterbalanced by an inertial force (mass times acceleration). Consider, for example, a microscopic swimmer that moves by changing its shape. Clearly the sum of all internal forces must be zero, i.e., the organism cannot "bootstrap" its own motion. However, changing its shape elicits reactive resistive forces from the fluid which themselves have to sum to zero. It turns out that the sequence of shape changes is uniquely determined by the requirement that the resistive forces cancel and can result in net motion of the swimmer. The requirement that there is net motion then constrains the allowed sequence of motions, as illustrated by the so-called scallop theorem formulated by Purcell [517]. Consider a scallop that moves in water at high Re by slowly opening and rapidly closing its shell. The latter action expels a jet of water that propels the scallop in the opposite direction, whereas the drag associated by reopening can be reduced by opening slowly. In contrast, at low Reynolds number, the flow of water into and out of the scallop over one cycle would be the same, regardless of the speed, implying that a scallop would make no net progress at low Re. This theorem reflects the fact that the Navier–Stokes equation in the limit Re$\to 0$ is time reversal symmetric, that is, it doesn't change under the transformations $t \to -t$ and $\mathbf{v} \to -\mathbf{v}$. The motion is independent of the speed and is determined by the sequence of body configurations. Since the opening and closing of a scallop's shell is time reversible, there is no net progress at low Re. The helical structure of flagella motors clearly breaks time reversal symmetry and thus allows *E. coli* to generate net motion at low Re.

How to swim at low Re. We now give a more abstract mathematical description of how to achieve net motion at low Re due to Shapere and

Wilczek [583] (see also [122]). Consider a cyclic sequence of shape changes of a microscopic swimmer at low Re in a Newtonian fluid. Let $S(t)$ denote the shape of the swimmer at time t at the actual orientation and location in space. $S(t)$ can be decomposed in terms of a displacement and orientation operator $\mathscr{R}(t)$ acting on a shape function $S_0(t)$, where $\{S_0\}$ denotes the set of all possible shapes at a fixed location and orientation. This is illustrated in Fig. 5.3, which shows two shapes $S_0(0)$ and $S_0(t)$ determined by a fixed local coordinate system (x,y). The actual physically located shape $S(t)$ is obtained by displacing and rotating the local coordinate system by the rigid body transformation $\mathscr{R}(t)$. For example, suppose that $S(\sigma)$ is a simply connected shape in $\mathbb{R}^3$, which is treated as a map from the two-sphere S^2 to $\mathbb{R}^3$ with $\sigma \in S^2$. Then $\mathscr{R}$ can be represented as a 4×4 matrix

$$\mathscr{R} = \begin{pmatrix} \mathbf{R} & \mathbf{d} \\ 0 & 1 \end{pmatrix},$$

where $\mathbf{R}$ is a standard 3×3 rotation matrix and $\mathbf{d} = (d_1, d_2, d_3)^T$ represents displacements of the global Cartesian coordinates (X,Y,Z). The operator $\mathscr{R}$ operates on the 4-vector $(\mathbf{S}_0(\sigma), 1)^T$ with $S_0(\sigma) \in \mathbb{R}^3$. It follows that under the rigid body transformation, $\mathbf{S}_0(\sigma) \to \mathbf{R}\mathbf{S}_0(\sigma) + \mathbf{d}$. It should be noted that the physical shape $S(t)$ is independent of the choice of local coordinates used to define the shapes S_0—changing the local coordinate system changes $\mathscr{R}(t)$ and hence $\mathbf{A}(t)$, but the sequence of motions is invariant under these transformations. Introduce a matrix $\mathbf{A}(t)$ for infinitesimal motions according to

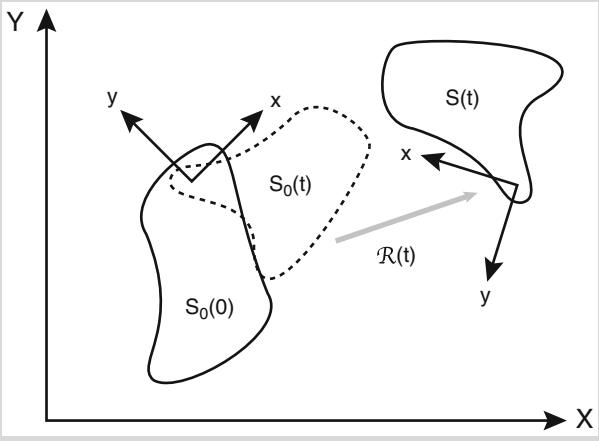


Fig. 5.3: Representing a shape $S(t)$ in physical space (X,Y) by a shape $S_0(t)$ at a fixed location and orientation, which is then shifted and reoriented by a rigid body transformation $\mathscr{R}(t)$

$$\frac{d\mathscr{R}}{dt} = \mathscr{R}(t)\mathbf{A}(t).$$

This can be formally integrated to give

$$\mathscr{R}(t) = \mathbb{T}\exp\left(\int_0^t \mathbf{A}(t')dt'\right).$$

Here $\mathbb{T}$ denotes the time-ordering operator. That is, on Taylor expanding the exponential, we obtain a series of multiple integrals involving products of the operators $\mathbf{A}(t)$ at different times. The time-ordered product means that operators at later times appear to the right of operators at earlier times—in general the operators don't commute:

$$\mathbb{T}\exp\left(\int_0^t \mathbf{A}(t')dt'\right) = 1 + \int_0^t \mathbf{A}(t')dt' + \int_0^t \int_0^{t''} \mathbf{A}(t')\mathbf{A}(t'')dt'dt'' + \ldots.$$

Finally, one can express $\mathbf{A}(t)$ in a time-independent manner by setting $\mathbf{A}(t)dt \equiv \mathbf{A}[S_0(t)]dS_0$ such that

$$\mathscr{R}(t) = \mathbb{T}\exp\left(\int_{S_0(0)}^{S_0(t)} \mathbf{A}[S_0]dS_0\right).$$

In the case of a cyclic sequence of shape changes, the net rotation and displacement per cycle period Δ is

$$\mathscr{R}(\Delta) = \mathbb{T}\exp\left(\oint \mathbf{A}[S_0]dS_0\right).$$

Of course, in order to calculate this explicitly one still has to solve the fluid dynamics equations at low Re to determine $A[S_0]$. Various examples can be found in [583].

5.3.1 Receptor Clustering and Signal Amplification

The main components of the signaling transduction pathways involved in *E. coli* chemotaxis are shown in Fig. 5.4 [628, 645]. Each chemoreceptor forms a complex with kinase CheA via an adaptor protein CheW—a protein kinase is an enzyme that modifies other proteins by chemically adding phosphate groups to them (phosphorylation). The autophosphorylation (self-activation) of CheA is suppressed (enhanced) when a chemoattractant (chemorepellant) binds to the associated receptor. In the activated state, CheA transfers a phosphate group to the motor regulator CheY thus counteracting dephosphorylation by CheZ. The phosphorylated form of CheY then diffuses away and binds with a flagellar motor, which then increases the motor's

clockwise bias and hence the cell's probability of tumbling. As with other biological sensory systems, the bacterial chemotaxis pathway allows the cell to adapt to persistent chemical stimuli. Adaptation is mediated by the methylation and demethylation of the chemoreceptors by the enzymes CheR and CheB*, where CheB* is the phosphorylated form of CheB that is also targeted by the activated form of CheA (see Sect. 5.3.2).

There are more than 10,000 chemoreceptors in a single *E. coli* cell, and they tend to form clusters around the cell pole. As first hypothesized by Bray et al. [62] and later confirmed experimentally [436, 605], one important function of receptor clustering is signal amplification due to cooperativity, analogous to cooperativity between multiple binding sites on a ligand-gated ion channel (see Sect. 3.1). Let us first consider a single chemoreceptor, which can bind to a single ligand molecule. Suppose that the kinase activity of the chemoreceptor (via CheA) has two discrete states, active (a) and inactive (i), and that the equilibrium constants for ligand binding/unbinding are different in the two states. We thus have a version of the MWC model with $n = 1$. It immediately follows (see equation (3.1.13)) that the steady-state probability of being in the active state is

$$p_{\rm a} = \frac{Y_0(1+K_a[L])}{Y_0(1+K_a[L])+(1+K_i[L])}, \tag{5.3.3}$$

where Y_0 is the equilibrium constant for $i \rightleftharpoons a$ and K_i, K_a are the equilibrium constants for ligand binding in the inactive and active states, respectively, with $K_a < K_i$. From equilibrium thermodynamics we know that $Y_0 = {\rm e}^{-\Delta E/k_B T}$, where $\Delta E = E_a - E_i$ is the free energy difference between the active and inactive states in the absence of a ligand. Given equation (5.3.3), we can also define an effective free energy difference $\Delta\overline{E}$ between the two states that is "averaged" with respect to the binding state by setting

$$p_{\rm a} = \frac{1}{1+{\rm e}^{\Delta\overline{E}/k_B T}}.$$

A comparison with (5.3.3) shows that

$$\Delta\overline{E}([L]) = \Delta E + k_B T \ln\frac{1+K_i[L]}{1+K_a[L]}. \tag{5.3.4}$$

Since $K_i > K_a$, increasing the ligand concentration $[L]$ increases the effective free energy and thus decreases the probability of being in the active state.

There are two basic models of receptor clustering. One involves dividing the receptors into a set of independent subclusters. Within each subcluster, all the receptors are tightly coupled and always in the same state (active or inactive), which is uncorrelated with the collective state of any other subcluster. However, the binding state of each receptor within a subcluster varies independently as in the single receptor case. This all-or-none activation state of a subcluster is precisely the MWC model introduced in Sect. 3.1. In the case of N receptors in a subcluster we obtain an equation of the form (3.1.13) for the probability p_a. Another way to derive such an equation is to note that the effective free energy for N globally coupled receptors is $E_N = N\Delta\overline{E}$ and $p_a = (1+{\rm e}^{E_N/k_B T})^{-1}$. On using (5.3.4), we have

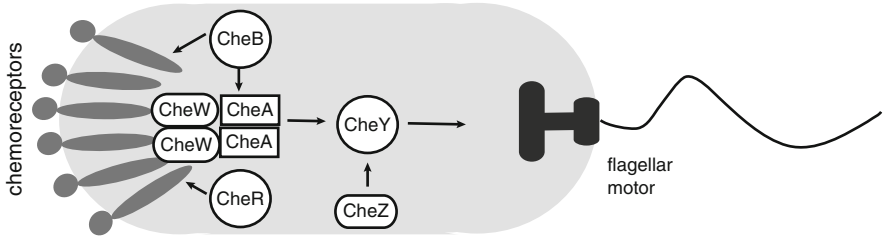


Fig. 5.4: Schematic diagram of major signaling pathways in *E. coli* chemotaxis. See text for details

$$p_{\mathrm{a}} = \left[1+\exp\left(N\left(\Delta\overline{E}+k_BT\ln\frac{1+K_i[L]}{1+K_a[L]}\right)\right)\right]^{-1} = \frac{Y_0(1+K_a[L])^N}{Y_0(1+K_a[L])^N+(1+K_i[L])^N}. \tag{5.3.5}$$

If $K_{\mathrm{a}} \ll [L]^{-1} \ll K_{\mathrm{i}}$, then we obtain a Hill function of order n:

$$p_{\mathrm{a}} = \frac{Y_0}{Y_0+K^n[L]^n} \tag{5.3.6}$$

where $K = K_{\mathrm{i}}$. It follows that

$$\frac{dp_{\mathrm{a}}}{d[L]} = -n[L]^{n-1}\frac{Y_0}{(Y_0+K^n[L]^n)^2} = -\frac{n}{[L]}p_{\mathrm{a}}(1-p_{\mathrm{a}}).$$

This suggests that maximal sensitivity will occur if the system is kept in a regime where $p_{\mathrm{a}} \approx 0.5$.

An alternative model of receptor clustering is to take the receptors to be distributed on some form of lattice with nearest neighbor interactions [151]. Let $m = 1,\ldots,N$ be a receptor label and denote the state of the mth receptor by a_m with $a_m = 1$ (active) or $a_m = 0$ (inactive). Let $\mathbf{a} = (a_1,\ldots,a_N)$ denote a given cluster state and take the corresponding free energy to be

$$H(\mathbf{a}) = -J\sum_{\langle m,n\rangle}(2a_m-1)(2a_n-1)+\Delta\overline{E}([L])\sum_m a_m. \tag{5.3.7}$$

The first term on the right-hand side represents interactions between nearest neighbor pairs on the lattice, denoted by $\langle n,m\rangle$, with J, $J>0$, a coupling strength. Such coupling favors neighboring receptors to be in the same activation state (1 or 0). The second term takes into account the internal energetics of individual receptors. In steady state the probability $P(\mathbf{a})$ of the cluster state $\mathbf{a}$ is given by

$$P(\mathbf{a}) = \frac{1}{Z}\mathrm{e}^{-H(\mathbf{a})/k_BT}, \quad Z = \sum_{\mathbf{a}}\mathrm{e}^{-H(\mathbf{a})/k_BT}.$$

The second model of receptor clustering is identical in structure to the Ising model of a 1D polymer considered in example 4.3 of Sect. 4.5. The latter is obtained under the transformations $2a_m - 1 \to \sigma_m = \pm 1$ and $\Delta\overline{E} \to -\alpha$, where σ_m is the orientation of the mth polymer link and α represents an external force. An exact solution for the mean number of active receptors can then be obtained along identical lines to the derivation of the force–extension relation for the polymer (see Ex. 5.2):

$$a \equiv \frac{1}{N}\left\langle \sum_m a_m \right\rangle = \frac{1}{2}\left[1 - \frac{\sinh\Delta\overline{E}([L])}{\sqrt{\sinh^2\Delta\overline{E}([L]) + \mathrm{e}^{-4J}}}\right]. \tag{5.3.8}$$

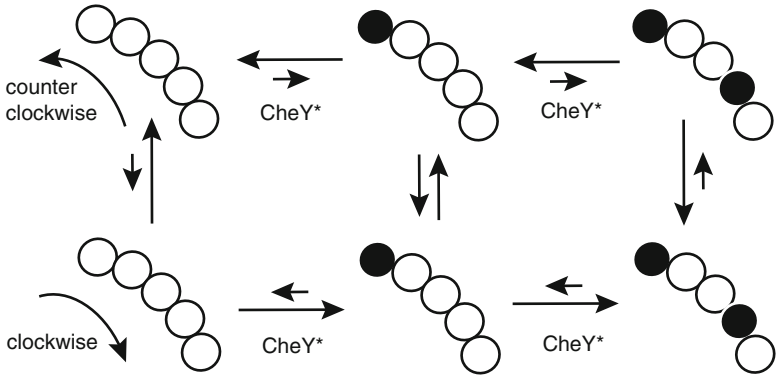


Fig. 5.5: Model of motor as a sensor of CheY. See text for details. Redrawn from [44]

It turns out that the flagellar motors are very sensitive to changes in the concentration of the phosphorylated form of CheY. Since ligand binding to chemoreceptors reduces the level of phosphorylation, this provides another mechanism for signal amplification. In Fig. 5.5, we show a model for the modulation of motor rotation bias due to binding of phosphorylated CheY (denoted by CheY*) [44]. Each flagellar motor is a rotary engine with a ring-like structure. The CheY molecules bind independently to multiple sites distributed around the ring with the binding affinity greater when the motor is rotating clockwise (CW) rather than CCW, that is, the associated equilibrium constants satisfy $K_{\mathrm{CW}} > K_{\mathrm{CCW}}$. When all sites are empty the equilibrium is biased towards CCW rotation. However, as more sites become occupied the equilibrium shifts to CW rotation. It follows that increasing the concentration of CheY* will favor the latter state. From the perspective of binding reactions, this model is also identical to the MWC model of a ligand-gated ion channel with n binding sites (Sect. 4.1). That is, we can map the open (R) and closed (T) states of the ion channel to the CW and CCW states of the molecular motor. Taking Y_0 to be the equilibrium constant for the switching between the CW and CCW states with all

sites empty, we see that the probability P_{CW} that the motor is in the CW state is (see equation (3.1.13))

$$P_{\mathrm{CW}} = \frac{Y_0(1+K_{\mathrm{CW}}c)^n}{Y_0(1+K_{\mathrm{CW}})^n + (1+K_{\mathrm{CCW}}c)^n} \tag{5.3.9}$$

where c denotes the concentration of CheY*.

5.3.2 Adaptation in Signal Transduction Pathways

A sudden increase in the concentration of a chemoattractant results in a decrease in the cell's tumbling frequency, but over a longer time scale the frequency recovers to its prestimulus level [39]. This frequency adaptation occurs over a wide range of stimulus strengths. At the molecular level, adaptation is mediated by the enzymes CheR and CheB*, which are responsible for the methylation and demethylation of chemoreceptors. The level of methylation ν, say, can be incorporated into the receptor clustering models by taking the difference in free energies between the active and inactive receptor states to depend on ν, that is, $\Delta E = \Delta E(\nu)$. As highlighted by Barkai and Leibler [20], the level of parameter fine-tuning that would be needed to account for the observed adaptation is unrealistic given the presence of noise. Therefore, they proposed a robust adaptation mechanism that doesn't need any fine-tuning. The basic idea is to assume that the relatively slow process of demethylation (CheB*) counteracts the shift in tumbling frequency induced by changes in the level of kinase activity (CheA*). More specifically, the Barkai–Leibler model assumes the following: (i) the rate of catalysis of the methylation enzyme CheR is at its maximum so independent of any concentrations; (ii) the action of the demethylation enzyme CheB* on active receptors is given by a Hill function of index $n = 1$. Thus,

$$\frac{d\nu}{dt} = F(a) \equiv \Gamma_R - \frac{\Gamma_B a}{K_B + a}, \tag{5.3.10}$$

where $\Gamma_{R,B}$ are maximum catalytic rates and a is the average receptor activity. In general, this is a nonlinear equation for ν since a is a function of ν via its dependence on $\Delta E(\nu)$.

In order to illustrate how adaptation occurs, let us return to the MWC model. Equation (5.3.5) shows that the average receptor activity a is a function of the level of methylation ν according to

$$a = a([L], \nu) \equiv \left[1 + \exp\left(N\left(\Delta E(\nu) + k_B T \ln \frac{1+K_i[L]}{1+K_a[L]}\right)\right)\right]^{-1}. \tag{5.3.11}$$

Suppose that there is a fixed background ligand concentration $[L] = L_0$ and denote the equilibrium receptor activity by a_0. This must also correspond to the steady-state solution of (5.3.10) with $F(a_0) = 0$:

$$a_0 = \frac{K_B \Gamma_R}{\Gamma_B - \Gamma_R},$$

which is independent of ligand concentration. (The fixed point is globally stable as $F'(a) < 0$ and $da/d\nu > 0$.) Given this solution for a_0, equation (5.3.11) implies that $a_0 = a(\nu, L_0)$, which can be inverted to yield the steady-state methylation level $\nu_0 = \nu(a_0, L_0)$. Now suppose that there is a sudden change in ligand concentration, $[L] = L_1$. The average receptor activity rapidly changes to give the new equilibrium solution for ν_0 fixed,

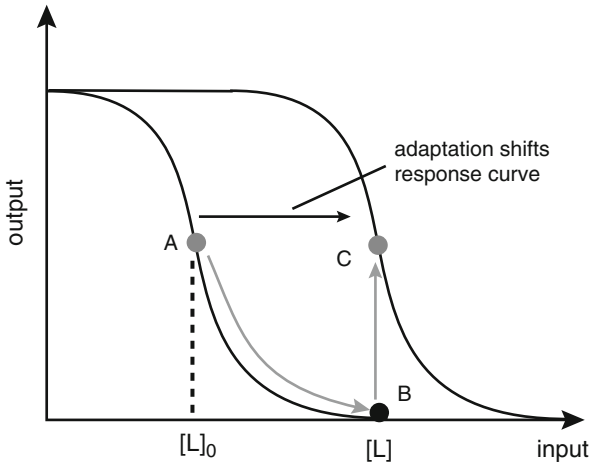


Fig. 5.6: Illustration of the Barkai–Leibler adaptation mechanisms. A sudden change in input (ligand concentration) $[L]_0 \to [L]$ induces a fast response from receptor activity state A to state B. Over longer time scales, the response curve shifts to a higher attractant concentration as the system adapts its methylation level until it reaches the adapted state C with the same activity as the prestimulus state A. As a result of the response curve shift, the high response sensitivity of adapted state C in the new environment is identical to that of state A. Redrawn from [645]

$$a_0 \to a_1 = a(L_1, \nu_0).$$

Suppose that $a_0 \approx 1/2$ so the system initially operates in the sensitive region of the response curve, whereas a_1 is outside this domain (see Fig. 5.6). However, over a longer time scale, the methylation level adapts according to equation (5.3.10) so that $a_1 \to a_0$ with $\nu_0 \to \nu_1 = \nu(a_0, L_1)$ (see Fig. 5.6). The linear response of the Barkai–Leibler model to a small oscillatory input is considered in Ex. 5.3.

5.3.3 Bacterial Chemotaxis as a Velocity-Jump Process

One of the challenges in modeling bacterial chemotaxis is understanding how extracellular biochemical signals are transduced into behavioral changes at the

macroscopic level illustrated in Fig. 5.2b. Early models tend to be phenomenological in nature, representing the dynamics of cells in terms of an advection–diffusion equation for the cell density $n(\mathbf{x},t)$, in which the velocity is taken to depend on the concentration gradient of some chemotactic substance [311, 326]. For example,

$$\frac{\partial n}{\partial t} = \nabla \cdot (D\nabla n - n\chi(c)\nabla), \tag{5.3.12}$$

where c is the concentration of the extracellular signal and the function $\chi(c)$ is known as a sensitivity function. Often the above equation is coupled to a reaction–diffusion equation for the evolution of c, which may itself depend on n if cells secrete their own chemoattractant. An alternative, stochastic formulation of bacterial motion has been developed in terms of a so-called *velocity jump process*, in which the velocity of the cell can randomly jump according to a discrete or continuous Markov process [6, 140, 262, 486, 487]. For example, let $p(\mathbf{x},\mathbf{v},t)$ denote the probability density of cells at position $\mathbf{x} \in \mathbb{R}^d$ and velocity $\mathbf{v} \in \mathbb{R}^d$ at time t. Then p evolves according to an equation of the form

$$\frac{\partial}{\partial t}p(\mathbf{x},\mathbf{v},t) + \mathbf{v}\cdot\nabla p(\mathbf{x},\mathbf{v},t) = -\lambda p(\mathbf{x},\mathbf{v},t) + \lambda \int T(\mathbf{v},\mathbf{v}')p(\mathbf{x},\mathbf{v}',t)d\mathbf{v}'. \tag{5.3.13}$$

Here λ is a constant turning rate, with $1/\lambda$ measuring the mean run length between velocity jumps. For simplicity the time spent in the tumbling state is neglected. The kernel $T(\mathbf{v},\mathbf{v}')$ is the conditional probability of a velocity jump from $\mathbf{v}'$ to $\mathbf{v}$ given that a jump occurs. If motion is restricted to 1D, then there are just two velocity states $\pm v$ and equation (5.3.13) reduces to the much simpler pair of equations

$$\frac{\partial p_+}{\partial t} + v\frac{\partial p_+}{\partial x} = -\lambda p_+ + \lambda p_-, \tag{5.3.14a}$$

$$\frac{\partial p_-}{\partial t} - v\frac{\partial p_-}{\partial x} = \lambda p_+ - \lambda p_-, \tag{5.3.14b}$$

where $p_\pm(x,t)$ are the probability densities of a cell being at (x,t) and moving to the right $(+)$ and left $(-)$, respectively, and v is the speed. This pair of equations is identical in form to the Dogterom–Leibler model of microtubule catastrophe [146] introduced in Sect. 4.1 [see equation (4.1.10)]. For constant v and λ, it can be reduced to a damped wave equation for the total density $p = p_+ + p_-$ known as the telegraph equation [223, 311, 487] (see Ex. 5.4).

In order to model 1D chemotaxis, it is necessary to introduce some bias into the stochastic switching (tumbling) between the velocity states $\pm v$ that depends on the extracellular concentration gradient c. One phenomenological way to achieve this is to assume that the rate of tumbling depends on the time derivative of the concentration $c(t) = c(x(t))$ along the bacterial trajectory according to some function $r(\dot{c})$, where $\dot{c} = \pm v dc/dx$ [44]. This yields the pair of equations

$$\frac{\partial p_+}{\partial t} + v\frac{\partial p_+}{\partial x} = -\frac{1}{2}r(vc'(x))p_+(x,t) + \frac{1}{2}r(-vc'(x))p_-(x,t) \tag{5.3.15a}$$

$$\frac{\partial p_-}{\partial t} - v\frac{\partial p_-}{\partial x} = \frac{1}{2}r(vc'(x))p_+(x,t) - \frac{1}{2}r(-vc'(x))p_-(x,t). \tag{5.3.15b}$$

We are assuming that when the bacterium tumbles there is an equal probability of moving in either direction and that tumbling is instantaneous—experimentally it is an order of magnitude faster than a typical run length. Another simplification is to take the tumble rate to depend on instantaneous values of the concentration gradient rather than a time-averaged change in concentration. The steady-state probability densities satisfy the pair of equations

$$v\frac{\partial p_+}{\partial x} = \frac{1}{2}r(-vc'(x))p_-(x) - \frac{1}{2}r(vc'(x))p_+(x)$$

and

$$-v\frac{\partial p_-}{\partial x} = \frac{1}{2}r(vc'(x))p_+(x) - \frac{1}{2}r(-vc'(x))p_-(x).$$

Adding these two equations gives

$$v\frac{\partial p_+}{\partial x} - v\frac{\partial p_-}{\partial x} = 0,$$

which implies that the difference $p_+(x) - p_-(x) = \text{constant}$. Assuming that $-\infty < x < \infty$, normalizability of the probability densities requires this constant to be zero. Hence, $p_\pm(x) = p(x)/2$ with $p(x)$ satisfying the single equation

$$v\frac{\partial p}{\partial x} = \frac{1}{2}\left[r(-vc'(x)) - r(vc'(x))\right]p(x).$$

Under the linear approximation $r(z) \approx r(0) + r'(0)z$, we have

$$r(\pm vc'(x)) \approx r(0) \pm r'(0)vc'(x),$$

and

$$v\frac{\partial p}{\partial x} = -r'(0)c'(x)p(x).$$

This has the straightforward solution

$$p(x) = \frac{1}{Z}\mathrm{e}^{-r'(0)c(x)}, \tag{5.3.16}$$

where Z is a normalization factor. If the signaling molecules correspond to a chemoattractant, then the rate of tumbling decreases in the direction for which $\dot{c} > 0$, that is, $r'(0) < 0$, and maxima of the steady-state solution (5.3.16) coincide with maxima of the concentration $c(x)$. Conversely, $r'(0) > 0$ for a chemorepellant and maxima of $P(x)$ coincide with minima of the concentration.

Erban and Othmer [170] have developed a more detailed 1D model of chemotaxis that incorporates aspects of the biochemical signal transduction pathways described in earlier parts of this section. For simplicity, they assume that there are two internal state variables of a bacterium, y_1, y_2, which evolve according to the equations

$$\tau_1 \frac{dy_1(t)}{dt} = g(c(x(t)) - y_1(t) - y_2(t), \quad \tau_2 \frac{dy_2(t)}{dt} = g(c(x(t)) - y_2(t). \tag{5.3.17}$$

Here $x(t)$ is the current position of a cell and the function g represents signal amplification based on a sigmoid, say, $g(c) = c/(c + K_d)$. For a constant signal $c = c_0$, we have the asymptotic solutions

$$\lim_{t\to\infty} y_1(t) = 0, \quad \lim_{t\to\infty} y_2(t) = g(c_0),$$

which implies that the variable y_1 adapts perfectly to a constant background signal. Finally, the turning rate is taken to be a linear function of y_1, $\lambda = \lambda_0 - \beta y_1$, where λ_0 is the basal switching frequency for a fully adapted cell and β is a positive constant. As a further simplification, suppose that $\tau_1 = 0$ and $g(c) = c$ so that $y_1(t) = c(x(t)) - y_2(t) \equiv -z(t)$. It follows that $\lambda = \lambda_0 + \beta z$ and

$$\tau_2 \frac{dz}{dt} = c(x(t)) - y_2(t) - \tau_2 c'(x(t)) \frac{dx}{dt} = -z(t) \mp \tau_2 c'(x(t))v,$$

depending on the sign of the velocity. Now let $p_\pm(x,z,t)$ denote the probability density for being at position x and in internal state z at time t with velocity $\pm v$. The resulting 1D model of chemotaxis takes the form [170]

$$\frac{\partial p_+}{\partial t} + v\frac{\partial p_+}{\partial x} + \frac{\partial}{\partial z}\left[\left(-\frac{z}{\tau_2} - vc'(x)\right)p_+\right] = (\lambda_0 + \beta z)[-p_+ + p_-] \tag{5.3.18a}$$

$$\frac{\partial p_-}{\partial t} - v\frac{\partial p_-}{\partial x} + \frac{\partial}{\partial z}\left[\left(-\frac{z}{\tau_2} + vc'(x)\right)p_-\right] = (\lambda_0 + \beta z)[p_+ - p_-]. \tag{5.3.18b}$$

Equations (5.3.18) can be analyzed by constructing moment equations and carrying out an appropriate truncation in order to solve the resulting closure problem [170]. First, introduce the following macroscopic variables:

$$n(x,t) = \int_{\mathbb{R}} [p_+(x,z,t) + p_-(x,z,t)]dz \tag{5.3.19a}$$

$$j(x,t) = v\int_{\mathbb{R}} [p_+(x,z,t) - p_-(x,z,t)]dz \tag{5.3.19b}$$

$$n_1(x,t) = \int_{\mathbb{R}} z[p_+(x,z,t) + p_-(x,z,t)]dz \tag{5.3.19c}$$

$$j_1(x,t) = v\int_{\mathbb{R}} z[p_+(x,z,t) - p_-(x,z,t)]dz \tag{5.3.19d}$$

$$j_2(x,t) = v\int_{\mathbb{R}} z^2[p_+(x,z,t) - p_-(x,z,t)]dz. \tag{5.3.19e}$$

Here $n(x,t)$ is the total probability density (summed over internal states and the two velocity states), $j(x,t)$ is the macroscopic particle flux, and the other macroscopic variables involve first and second moments with respect to the slow internal variable z. Multiplying equations (5.3.18) by 1 or z, integrating with respect to z, and adding or subtracting the resulting equations leads to the following hierarchy of moment equations (see Ex. 5.5):

$$\frac{\partial n}{\partial t}+\frac{\partial j}{\partial x}=0, \tag{5.3.20a}$$

$$\frac{\partial j}{\partial t}+v^2\frac{\partial n}{\partial x}=-2\lambda_0 j-2\beta j_1, \tag{5.3.20b}$$

$$\frac{\partial n_1}{\partial t}+\frac{\partial j_1}{\partial x}=-c'(x)j-\frac{n_1}{\tau_2}, \tag{5.3.20c}$$

$$\frac{\partial j_1}{\partial t}+v^2\frac{\partial n_1}{\partial x}=-v^2c'(x)n-(2\lambda_0+\tau_2^{-1})j_1-2\beta j_2. \tag{5.3.20d}$$

As they stand, these do not form a closed system of equations, since (n,n_1,j,j_1) couple to the second-order flux j_2, which will itself couple to higher-order moments. However, it is possible to achieve moment closure by assuming $j_2=0$. This can be justified rigorously provided that the concentration gradient is sufficiently shallow [170].

Equations (5.3.20) with $j_2=0$ can be used to determine various statistics of the motion. For the sake of illustration, suppose that the concentration gradient is a constant $c'(x)=c_0$. Assuming the normalization $\int_{\mathbb{R}} n(x,t)dx=1$, the mean position $\bar{x}$ and MSD σ^2 are defined according to

$$\bar{x}(t)=\int_{\mathbb{R}} xn(x,t)dx, \quad \sigma^2(t)=\int_{\mathbb{R}}(x-\bar{x})^2n(x,t)dx. \tag{5.3.21}$$

Multiplying equation (5.3.20a) by x or by $(x-\bar{x})^2$ and integrating with respect to x gives

$$\frac{d\bar{x}}{dt}=j_0, \quad \frac{d\sigma^2}{dt}=2j_1-2\bar{x}j_0, \tag{5.3.22}$$

where

$$j_s=\int_{\mathbb{R}} x^s j(x,t)dx, \quad j_{1s}=\int_{\mathbb{R}} x^s j_1(x,t)dx, \quad n_{1s}=\int_{\mathbb{R}} x^s n_1(x,t)dx. \tag{5.3.23}$$

A closed set of first-order differential equations for (j_0,n_{10},j_{10}) can be obtained by integrating equations (5.3.20b)–(5.3.20d) with respect to x. Similarly, a closed set of equations for the triplet (j_1,n_{11},j_{11}) can be derived by multiplying equations (5.3.20b)–(5.3.20d) by x and then integrating with respect to x. One finds that each of the two dynamical systems converges to a stable fixed point so that in the limit $t\to\infty$ [170],

$$\bar{x}(t) = \frac{\beta v^2 c_0 \tau_2}{\lambda_0 + 2\lambda_0^2 \tau_2} t, \quad \sigma^2(t) = \left(\frac{v^2}{\lambda_0} + \frac{2\beta v^4 c_0^2 \tau_2^3}{(\lambda_0 + 2\lambda_0^2 \tau_2)^2} \right) t. \tag{5.3.24}$$

The details of the calculation are left to Ex. 5.5.

In summary, the large-time behavior of a bacterium undergoing chemotaxis in a shallow concentration gradient is characterized by constant drift and diffusion. This is a special case of a more general result that, under an appropriate rescaling of space and time, wave-like (hyperbolic) equations for chemotaxis such as the general velocity-jump process of equation (5.3.13) can be approximated in the long-time limit by an advection–diffusion (parabolic) equation, provided that the concentration gradient is sufficiently shallow [262, 486]. We will encounter analogous results for PDE models of motor-driven transport in Sect. 7.4.

5.4 Hair Cells and Active Mechanotransduction

Most sensory cells have to amplify their signals in order to separate them from background noise (see also Sect. 5.1). For example, photoreceptors enhance their responses to photon excitation a thousandfold by using a biochemical cascade [44]. On the other hand, hair cells of the vertebrate inner ear use a mechanical active process to amplify their inputs [286, 412, 422]. When sound reaches the cochlea—a spiraled, hollow, conical chamber of bone in the inner ear along which waves propagate—it elicits mechanical vibrations that stimulate hair cell receptors. These receptors transduce the vibrations into an electrical signal via mechanotransduction, simultaneously performing work that amplifies the mechanical signal resulting in positive feedback. The hair cells of all vertebrates share a similar structure and transduce mechanical stimuli according to the same basic mechanism [287]. On the top of each hair cell is a cluster of 20–300 actin-based cylindrical structures called *stereocilia*, which is known as the *hair bundle* (see Fig. 5.7). The stereocilia develop

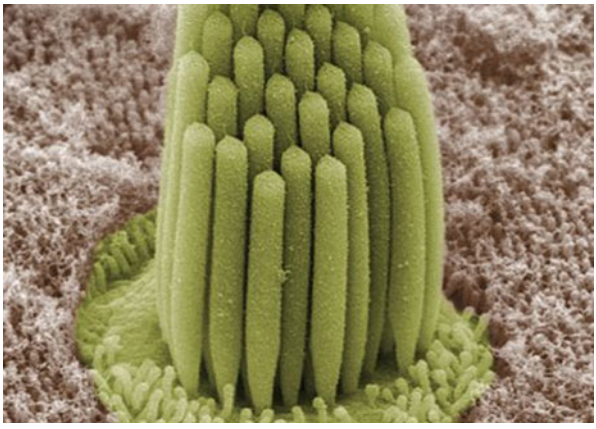


Fig. 5.7: Electron micrograph showing stereocilia of an inner hair cell of the bullfrog (Public domain figure downloaded from Wikipedia Commons)

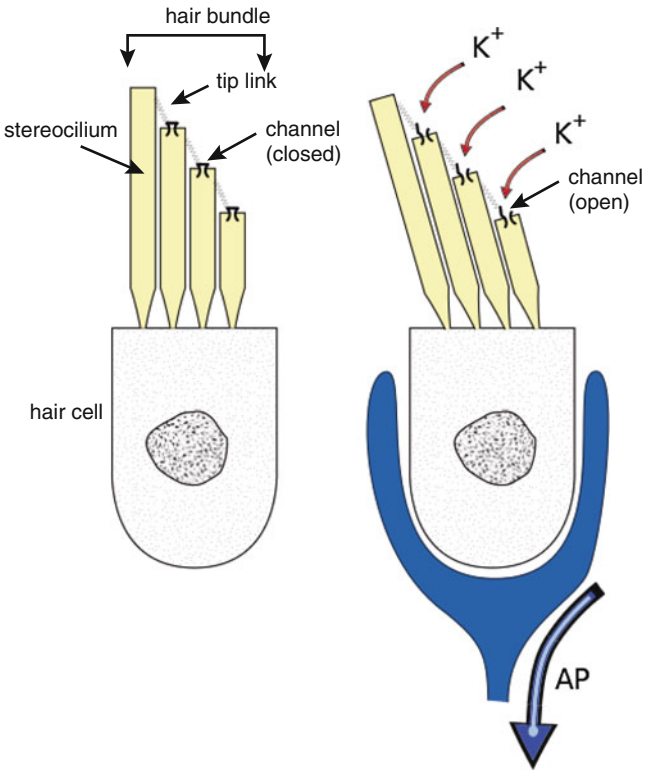


Fig. 5.8: Schematic illustration of a hair cell, illustrating how a mechanical stimulus deflects the bundle of stereocilia surmounting the cell resulting in the opening of mechanosensitive ion channels and the influx of K^+ and subsequent influx of Ca^{2+}. This can lead to the firing of an action potential (AP) (Public domain figure downloaded from Wikipedia Commons)

in such a way that there is a specific variation of their lengths across the hair bundle giving the latter a beveled shape (see Fig. 5.8). The mechanical stimulus induced by sound reaching the ear deflects the hair bundles, with their component stereocilia bending at their base. This deflection causes a shearing motion between neighboring stereocilia, which is detected by mechanosensitive ion channels located near the stereociliary tips. This transduction is mediated by cadherin-based adhesive tip links that couple adjacent stereocilia and can open the ion channels under tension. This allows K^+ to enter the hair cell. The resulting depolarization opens voltage-gated Ca^{2+} channels and the intracellular Ca^{2+} concentration rises. This in turn opens Ca^{2+}-sensitive K^+ channels through which K^+ can flow out of the cell and the cell returns to rest. The given sequence of events can result in an action potential being produced by the hair cell.

One major feature of the hair bundle is that the tension of the tip link can be adjusted by myosin motors that walk up and down the stereocilia [607, 610], which then allows the hair cell to adapt to a sustained deflection of the hair bundle. It is thought that, at least in the case of nonmammalian tetrapods (four-legged vertebrates), the interaction of the molecular motors with the mechanical properties of the hair bundle forms the basis of active processes in the inner ear, which include signal amplification, enhanced frequency selectivity, and spontaneous oscillatory acoustic emissions. Moreover, from a dynamical systems perspective, these characteristics emerge naturally if the transduction process operates near a Hopf bifurcation [90, 109, 423, 465, 657]. In this section we review some of the models that have been developed to explore active processes in hair cells.

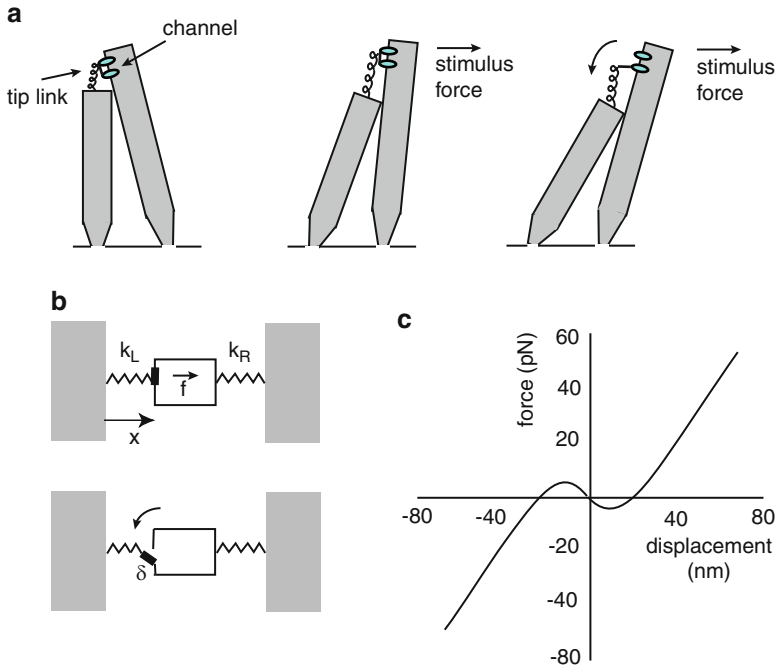


Fig. 5.9: Gating-spring model of mechanotransduction. (**a**) Schematic diagram of two stereocilia connected by a tip link that is attached to a transduction ion channel. Deflection of the bundle by a positively directed stimulus force bends the stereocilia, stretches the tip link, and consequently opens the ion channel, allowing K and Ca^{2+} to enter the cytoplasm and depolarize the hair cell. (**b**) Simplified mechanical model of coupled stereocilia. (**c**) Sketch of force–displacement curve similar to one found experimentally in hair cells of the bullfrog [423]. A region of negative spring stiffness can be seen

5.4.1 Gating-Spring Model

Our starting point is a gating-spring model of mechanotransduction in hair cells, as reviewed in [418]. Suppose that the hair bundle is modeled as a collection of N elastic units in parallel. Each unit consists of a mobile element of width ΔA that is attached to two fixed walls by a pair of springs in series (see Fig. 5.9b), which is a simplified model of the mechanical properties of a pair of linked stereocilia (see Fig. 5.9a). The left spring represents the tip link or gating spring (gs) and is attached to the element via a hinge or trapdoor, whereas the right spring (sp) represents the stereociliary pivots. Assume, for the moment, that the trapdoor is closed. Let x be the distance of the mobile element from the left-hand wall and denote the lengths of the gating spring and pivots by $a_{\rm gs}$ and $a_{\rm sp}$, respectively. If the distance between the walls is A, then

$$a_{\rm gs} = x, \quad a_{\rm sp} = A - \Delta A - x.$$

Suppose that each spring has an equilibrium length (no tension) denoted by $\overline{a}_{\rm gs}$ and $\overline{a}_{\rm sp}$. It follows that we can express the displacements form equilibrium of the two springs as

$$\Delta a_{\rm gs} = a_{\rm gs} - \overline{a}_{\rm gs} = x - \overline{x}_{\rm gs}, \quad \Delta a_{\rm sp} = a_{\rm sp} - \overline{a}_{\rm sp} = -(x - \overline{x}_{\rm sp}),$$

with $\overline{x}_{\rm gs} = \overline{a}_{\rm gs}$ and $\overline{x}_{\rm sp} = A - \Delta A - \overline{a}_{\rm sp}$. When the trapdoor is open, it is assumed that the left spring's length is reduced by an amount δ, so that $\Delta a_{\rm gs} \to \Delta a_{\rm gs} - \delta$. Now suppose that an external force f in the positive x-direction is applied to a single unit and that we have Hookean springs with spring constants $k_{\rm gs}$ and $k_{\rm sp}$, respectively. The displacement x is then determined by the force-balance equations

$$\begin{aligned} f = f_c(x) &\equiv k_{\rm gs}\Delta a_{\rm gs} - k_{\rm sp}\Delta a_{\rm sp} = k_{\rm gs}(x - \overline{x}_{\rm gs}) + k_{\rm sp}(x - \overline{x}_{\rm sp}) \\ &= K(x - \overline{x}), \end{aligned} \tag{5.4.1}$$

for the closed trapdoor and

$$\begin{aligned} f = f_o(x) &\equiv k_{\rm gs}(\Delta a_{\rm gs} - \delta) - k_{\rm sp}\Delta a_{\rm sp} = k_{\rm gs}(x - \overline{x}_{\rm gs} - \delta) + k_{\rm sp}(x - \overline{x}_{\rm sp}) \\ &= K(x - \overline{x}) - k_{\rm gs}\delta, \end{aligned} \tag{5.4.2}$$

for the open trap door, where $K = k_{\rm gs} + k_{\rm sp}$ and $\overline{x} = [k_{\rm gs}\overline{x}_{\rm gs} + k_{\rm sp}\overline{x}_{\rm sp}]/K$. (Note that the applied force is opposed by stretching the left spring and compressing the right spring.)

Having looked at the mechanical properties of the springs, we now have to incorporate the stochastic opening and closing of the trapdoor. This is achieved by treating the gating-spring unit as a two-state system in thermodynamic equilibrium, so that the probability of being in an open or closed state is given by a Boltzmann–Gibbs distribution (Sect. 1.4). That is $P_c(x) = 1 - P_o(x)$, with $P_o(x) = (1 + {\rm e}^{\Delta E(x)/k_BT})^{-1}$ and $\Delta E(x)$ is the energy difference between the open and closed states. There are two contributions to this energy difference. First, there is an increase ΔE_0 in the configuration energy of the trapdoor when it jumps to the

open state. Second, there is a change in the potential energy stored by the left spring when it shifts by an amount δ, which is given by $-\delta k_{\rm gs}(x-\bar{x}_{\rm gs})$. Hence,

$$P_o(x) = \frac{1}{1+{\rm e}^{-[\delta k_{\rm gs}(x-\bar{x}_{\rm gs})-\Delta E_0]/k_BT}}. \tag{5.4.3}$$

Now let us consider N identical gating-spring units in parallel. Since they are in parallel, each unit experiences the same applied force f/N. If N is sufficiently large, then fluctuations in the fraction of open and closed gates can be ignored. It follows that we have the force-balance equation

$$\begin{aligned} f = f(x) &\equiv N\left[f_o(x)P_o(x) + f_c(x)P_c(x)\right] \\ &= N\left[(K(x-\bar{x}) - k_{\rm gs}\delta)P_o(x) + K(x-\bar{x})P_c(x)\right] \\ &= N\left[K(x-\bar{x}) - k_{\rm gs}\delta P_o(x)\right] \\ &= N\left[K(x-\bar{x}) - \frac{\delta k_{\rm gs}}{1+{\rm e}^{-[\delta k_{\rm gs}(x-\bar{x}_{\rm gs})-\Delta E_0]/k_BT}}\right]. \\ &= K_{\rm tot}x + f_0 - \frac{Nz}{1+{\rm e}^{-z(x-x_0)/k_BT}}, \end{aligned} \tag{5.4.4}$$

where $K_{\rm tot} = NK$, $f_0 = -NK\bar{x}$, $z = \delta k_{\rm gs}$, and $x_0 = \bar{x}_{\rm gs} + \Delta E_0/z$. Fitting the model to experimental data from bullfrog hair cells yields the following example set of parameter values [422]: $N = 65$, $K_{\rm tot} = 10^3\ \mu{\rm Nm}^{-1}$, $z = 0.72$ pN, $f_0 = 25$ pN, and $x_0 = -2.2$ nm. If one plots the force–displacement function $f(x)$ for these values at room temperature, one obtains the curve shown in Fig. 5.9c. It can be seen that for sufficiently large displacements in the positive or negative x directions, the system acts like an ordinary spring, that is, it has approximately constant stiffness. However, within $\pm 20nm$ of the resting position, the stiffness df/dx varies significantly with displacement. Even more striking is that, within ± 10 nm of the resting position, the stiffness slope is negative and displacement of the bundle in a particular direction requires a force in the opposite direction—the hair bundle is said to have *channel compliance*. It is important to note that no active processes have been included in the model, since there is no net consumption of energy. However, when the force–displacement characteristics of the form shown in Fig. 5.9c are combined with the action of myosin motors, the active features of hair cells can be reproduced [422, 424] (see Sect. 5.4.2).

Before considering active processes, however, we briefly consider what happens when N is small so that fluctuations in the number of open ion channels cannot be ignored. From the analysis of Sect. 3.2, we know that the probability $P_N(n|x)$ of there being n open channels for a given displacement x is given by a binomial distribution of the form (3.2.8):

$$P_N(n|x) = P_o(x)^n(1-P_o(x))^{N-n}\frac{N!}{(N-n)!n!}.$$

It follows that the mean and variance of the number of open channels are

$$\langle n \rangle = NP_o(x), \quad \sigma^2 = NP_o(x)(1 - P_o(x)).$$

For a given displacement x, the necessary force on the hair bundle can be written as [see equation (5.4.4)],

$$f(x) = f_0(x) - zn,$$

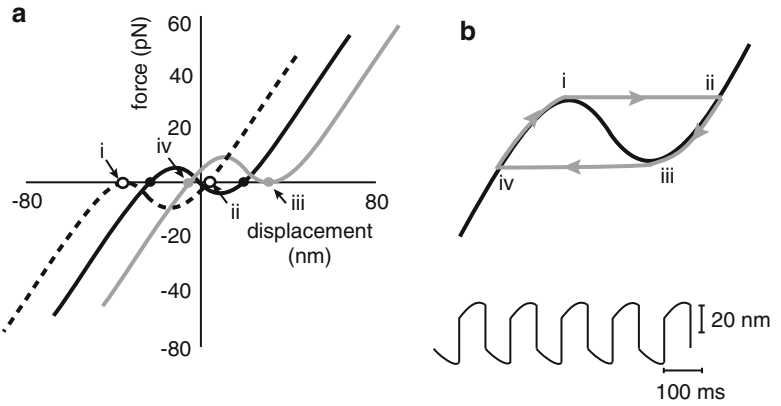


Fig. 5.10: Possible mechanism for spontaneous oscillations of a hair bundle. (**a**) Shifts in the force–displacement curve due to the action of myosin motors. See text for details. (**b**) Hair bundle acts like a relaxation oscillator with fast jumps ($i \to ii$ and $iii \to iv$) alternating with slow shifts in the fixed point due to adaptation ($ii \to iii$ and $iv \to i$). (**c**) Sketch of variation of displacement x with time (Redrawn from Martin et al. [422])

where $f_0(x) = NK(x - \bar{x})$ and n is the stochastic number of open ion channels. Thus $f_0(x)$ is the force needed to hold bundle at x when all the channels are closed. We can now determine the mean and variance of the force:

$$\langle f - f_0 \rangle = -z\langle n \rangle = -NzP_o(x),$$

and

$$\mathrm{Var}[f - f_0] = z^2\sigma^2 = Nz^2P_o(x)(1 - P_o(x)).$$

5.4.2 *Channel Compliance, Myosin Motors and Spontaneous Oscillations*

The channel compliance (negative stiffness) of the hair bundle implies that the bundle can operate in a bistable regime for sufficiently small applied forces. However, the force–displacement curve can be shifted by the Ca^{2+}-regulated action of myosin molecular motors that move up and down a stereocilium, altering the stiffness of the

tip link [422, 424]. The myosin motors adapt the response of the hair bundle so that if a negative (positive) displacement has been maintained for some time, the force–displacement curve of Fig. 5.9c is shifted to the left and downward (to the right and upward). This has the effect of moving the negative stiffness region toward the offset point. One possible mechanism for Ca^{2+} regulation of myosin is that Ca^{2+} simply reduces the probability that myosin motors bind to actin filaments, thus allowing the transduction element to lose tension [287]. Alternatively, Ca^{2+} could alter the equilibrium between different bound conformational states of the motor. It has also been suggested that Ca^{2+} might regulate hair bundle dynamics in a myosin-independent fashion [109, 287]. For example, the energy associated with binding of Ca^{2+} directly to the channel or to an associated protein could reduce the open probability. However, the myosin-dependent mechanism is currently thought to be more likely.

As shown by Martin et al. [422], the above adaptation mechanism can also result in spontaneous oscillations, which is illustrated in Fig. 5.10. Suppose that the bundle occupies a negative displacement equilibrium when $f = 0$ (black curve in Fig. 5.10a). The ion channels are then in a low open probability state and the Ca^{2+} concentration is kept at a low level by Ca^{2+} pumps. This upregulates the myosin motors, resulting in an increased stiffness of the tip link and a leftward shift of the force–displacement curve (dashed curve in Fig. 5.10a). A sufficient shift leads to the disappearance of the negative fixed point (i) and the system jumps to the corresponding positive fixed point (ii). The ion channels are now in a high open probability state, Ca^{2+} flows into the cell and downregulates the myosin motors, and the force–displacement curve shifts in a rightward direction. Eventually, the positive fixed point (iii) disappears (gray curve in Fig. 5.10a), and the system jumps back to a negative fixed point (iv). The cycle then repeats itself resulting in a periodic solution (see Figs. 5.10b,c). Since the jumps are much faster than adaptation, the hair bundle acts like a relaxation oscillator, analogous to the conductance-based models of a neuron considered in Sect. 3.5. (Indeed, it is possible to linearly transform equations (5.4.5) and (5.4.7) for fixed C to obtain equations similar in structure to those of the FitzHugh–Nagumo model [121].) It is thought that these oscillations are responsible for the spontaneous acoustic emissions observed in nonmammalian vertebrates. For example, Martin et al. [422] measured the power spectrum (see equation 2.2.5) of spontaneous oscillations emitted by the hair bundle of a bullfrog hair cell and found a sharp peak at around 8 Hz and a half-width of around 3 Hz; the spectral broadening is a result of thermal noise. The typical range of spontaneous oscillation frequencies is 5–50 Hz. The existence of spontaneous oscillations provides a possible mechanism for signal amplification and frequency tuning, namely, a nonlinear resonance effect when the stimulus frequency is sufficiently close to the natural frequency of spontaneous oscillations and the system operates close to a Hopf bifurcation point [90, 109, 158, 423, 465, 657].

For the sake of illustration, we will consider a dynamical model of spontaneous oscillations in hair bundles based on the combined action of Ca^{2+} and myosin motors on the gating-spring model [424, 465] (see also Ref. [657]). Neglecting

inertial effects, the position of a single gating-spring element evolves according to the equation

$$\xi \frac{dx}{dt} = -N\left[k_{\rm gs}(x - x_M - \delta P_0(x)) + k_{\rm sp}x\right] + F_{\rm ext} \tag{5.4.5}$$

where ξ is a friction coefficient, $F_{\rm ext}$ is an external force, and δ is scaled by a geometrical factor that takes into account the gain of stereociliary shear motion. The open probability is given by a slightly modified version of equation (5.4.3):

$$P_o(x) = \frac{1}{1 + A{\rm e}^{-[\delta k_{\rm gs}(x - x_M)]/k_BT}}, \quad A = {\rm e}^{[\Delta G + k_{\rm gs}\delta^2/(2N)]/k_BT}. \tag{5.4.6}$$

Following the idea that myosin motors modify the tension in the tip link of a channel, the effective equilibrium position x_M of the gating spring is maintained by N_M myosin motors pulling against the force $f(x) = k_{\rm gs}(x - x_M) - \delta P_0(x)$. That is, we identify x_M with the position of the motor cluster. The dynamics of the cluster is assumed to satisfy a linear force–velocity relation with slope ξ_M:

$$\xi_M \frac{dx_M}{dt} = Nk_{\rm gs}\left[x - x_M - \delta P_0(x)\right] - N_M f p(C). \tag{5.4.7}$$

Here the force exerted by the motors is taken to be proportional to the force f generated by a single motor and the probability p that a motor is bound to an active filament. The active force produced by the molecular motors corresponds to the motors climbing up the stereocilia, $dx_M/dt < 0$, which increases the tension of the gating springs and thus increases the open probability P_o of the ion channels. The binding probability $p = p(C)$ is assumed to be a monotonically decreasing function of the intracellular Ca^{2+} concentration C. Ignoring nonlinearities in $p(C)$, the binding probability can be written as $p(C) \approx p_0 - p_1 C$ with $p_{0,1} > 0$, provided that $C < p_0/p_1$. Finally, the intracellular Ca^{2+} dynamics is modeled as

$$\tau \frac{dC}{dt} = -(C - C_0) + [C_M - C_0]P_0(x), \tag{5.4.8}$$

where the decay term represents the effects of Ca^{2+} pumps and the other term on the right-hand side is the total flux through the open ion channels. When all the channels are closed, C returns to the background concentration C_0. A crucial aspect of the model is a separation of time scales—the channel kinetics are assumed to be much faster than the Ca^{2+} dynamics, which are themselves assumed to be faster than the bundle and motor dynamics.

Since $\tau \ll \xi, \xi_M$ (fast Ca^{2+} dynamics), one can use a slow–fast analysis and set $\tau = 0$. This yields an effective planar dynamical system for (x, x_M) given by equations (5.4.5) and (5.4.7) with $p(C) = p_0 - p_1 C_M P_o(x)$ for $C_0 = 0$. (Vilfan and Duke [657] also use a slow–fast decomposition but treat the motors as the slow system by fixing x_m and consider the planar dynamics of (x, C).) One can then determine the existence and stability of fixed points for $F_{\rm ext} = 0$ and derive conditions for the occurrence of a Hopf bifurcation along the lines of Box 3B. Nadrowski et al. [465] constructed a bifurcation diagram in terms of two parameters: the maximal force

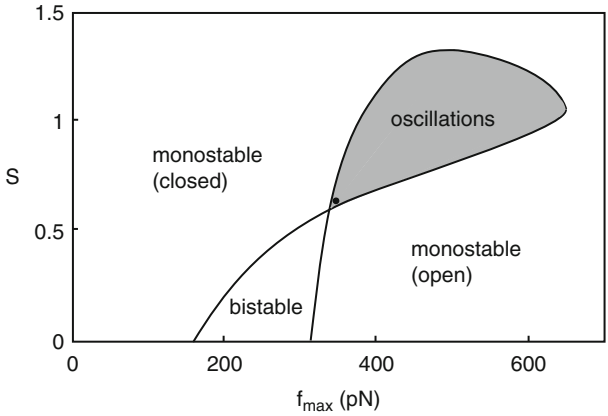


Fig. 5.11: Bifurcation diagram of hair bundle model given by equations (5.4.5)–(5.4.8) for $\tau = 0$. Different dynamical regimes are shown as a function of the maximal motor force $f_{\max}$ and the strength S of Ca^{2+} feedback on motor activity. There are two monostable domains where either most of the ion channels are closed or most are open. The *shaded region* indicates where stable oscillations exist, and Hopf bifurcations occur along the boundary of this region. For smaller $f_{\max}$ (towards the bistable region) the Hopf bifurcation is subcritical, whereas for larger $f_{\max}$ (away from the bistable regime) the Hopf bifurcation is supercritical. (The *black dot* indicates the point $f_{\max} = 50.3pn$, $S = 0.65$ where stochastic simulations produced the best quantitative agreement with experiments; see text for details.) Other parameter values are as follows: $\xi = 2.8\ \mu\text{Nsm}^{-1}$, $\xi_M = 10\ \mu\text{Nsm}^{-1}$, $N = 50$, $N_M = 3,000$, $\delta = 61$ nm, $k_{gs} = 15\ \mu\text{Nm}^{-1}$, $k_{sp} = 12\ \mu\text{Nm}^{-1}\ \mu\text{Nm}^{-1}$, $C_0 = 0$, $\Delta G = 10\ k_B T$, $T = 300$ K (Redrawn from Nadrowski et al. [465])

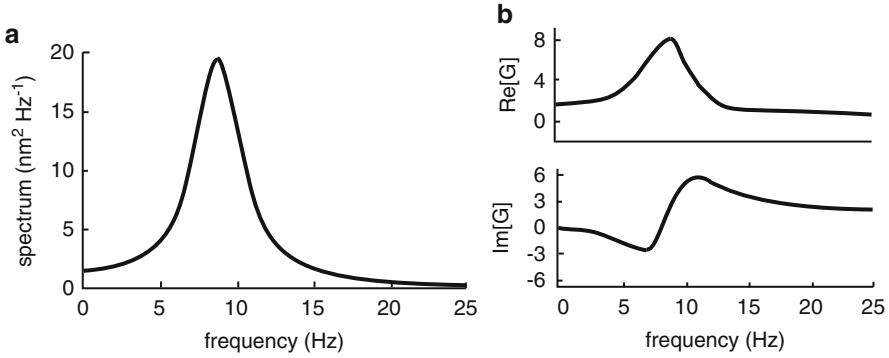


Fig. 5.12: Stochastic version of hair bundle model. (**a**) Power spectrum for spontaneous oscillations. (**b**) Real and imaginary parts of linear response function in response to sinusoidal forcing at an amplitude of 1pN. Parameter values given by *black dot* in Fig. 5.11 and $\tau_C = 1$ ms (Redrawn from Nadrwoski et al. [465])

$f_{\max} = \gamma N_m f p_0$ exerted by the molecular motors and the dimensionless strength $S = C_M p_1/p_0$ of the negative Ca^{2+} feedback. The final term on the right-hand side of (5.4.7) becomes $-f_{\max}(1 - SP_o(x))$. The bifurcation diagram is sketched in Fig. 5.11 and consists of different dynamical regimes as indicated. In particular spontaneous oscillations occur at intermediate values of the maximal force and the strength of Ca^{2+} feedback; in other regions the system is either in a monostable or bistable regime (see Ex. 5.6).

One simplification of the above model is that it ignores the effects of noise. Nadrowski et al. [465] also considered a stochastic version of the model by introducing white noise terms η, η_M, and η_C on the right-hand side of equations (5.4.5), (5.4.7), and (5.4.8), respectively. The Gaussian random variables are taken to have zero mean and autocorrelations

$$\langle \eta(t)\eta(0)\rangle = 2k_B T \xi \delta(t) \tag{5.4.9a}$$

$$\langle \eta_M(t)\eta_M(0)\rangle = 2k_B T_M \xi_M \delta(t) \tag{5.4.9b}$$

$$\langle \eta_C(t)\eta_C(0)\rangle = 2N^{-1} C_M^2 P_o(1 - P_o)\tau_C \delta(t). \tag{5.4.9c}$$

The major source of noise for the hair bundle is Brownian motion in the surrounding fluid, and one can use the Einstein relation to determine the noise strength in terms of the friction coefficient ξ. Although the motors also undergo Brownian motion, there are additional sources of noise due to the random binding and binding to filament tracks. This leads to an effective temperature $T_M \approx 1.5T$. Finally, the main source of noise for Ca^{2+} dynamics is the random opening and closing of ion channels, which can be described by a binomial distribution as outlined at the end of Sect. 5.4.1. Assuming the channel kinetics relaxation time τ_C is very fast, one can approximate the channel noise by white noise (see also Chap. 3). Simulations of the stochastic model for $\tau_C = 1ms$, $f_{\max} = 50.3pn$, and $S = 0.65$ (indicated by the black dot in the bifurcation diagram of Fig. 5.11) generates a spectrum of spontaneous oscillations that agrees quantitatively with experiments. An example of a spectrum obtained by Nadrowski et al. [465] is sketched in Fig. 5.12, together with the corresponding real and imaginary parts of the linear response function in the frequency domain.

5.4.3 *Active Amplification Close to a Hopf Bifurcation*

A number of theoretical studies have suggested that many of the active properties of a hair bundle can be reproduced by assuming that it operates close to a Hopf bifurcation [90, 109, 158, 422]. This is also consistent with the dynamical model of spontaneous oscillations considered by Nadrowski et al. [465] (see Fig. 5.11), although certain care needs to be taken since the Hopf bifurcation may be subcritical, so that there is a rapid transition to a large amplitude relaxation oscillator. Following Refs. [90, 158], we now consider the generic behavior of a forced oscillator close to a supercritical Hopf bifurcation. Let μ denote some bifurcation parameter of the system, which could be related to the activity of myosin motors or the concentration

of intracellular Ca^{2+} in the case of a hair bundle. Suppose that the system is in a stable stationary state for $\mu < 0$, whereas it exhibits spontaneous oscillations for $\mu > 0$ due to a supercritical Hopf bifurcation at the critical value $\mu_c = 0$. Recall from Box 3B that close to a Hopf bifurcation, the dynamics of an unforced oscillator with natural frequency ω_0 can be represented (after an appropriate change of variables) by the normal form (4.4.30):

$$\frac{dx}{dt} = \mu x + \omega_0 y - x(x^2+y^2), \quad \frac{dy}{dt} = -\omega_0 x + \mu y - y(x^2+y^2),$$

which can be recast in complex form by setting $z = x + iy$:

$$\frac{dz}{dt} = (\mu - i\omega_0)z - |z|^2 z. \tag{5.4.10}$$

Now suppose that we drive the oscillator with a forcing term $ae^{i\omega t}$, and look for solutions of the form $z = Ae^{i(\omega t+\theta)}$. Substituting into the complex version of the normal form gives

$$\left(iA(\omega - \omega_0) - \mu A + A^3\right) e^{i\theta} = a.$$

The relevant quantity in terms of amplification is the amplitude of the response, so taking the modulus of both sides we have

$$A^6 - 2\mu A + [\mu^2 + (\omega - \omega_0)^2]A^2 = a^2. \tag{5.4.11}$$

Solving this equation for fixed stimulus strength a, we can plot the amplitude A as a function of the stimulus frequency. The results are shown in Fig. 5.13 for various

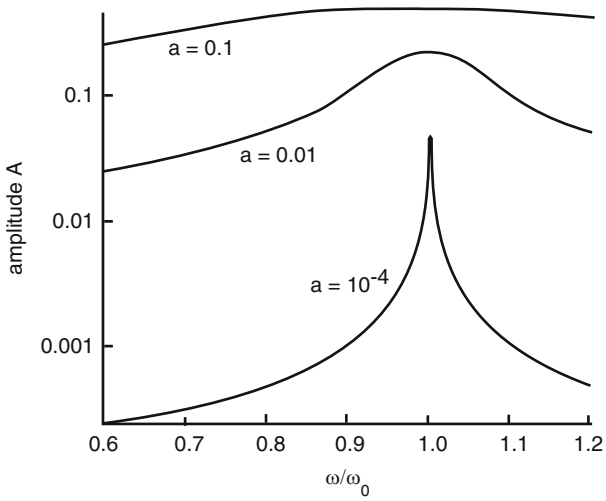


Fig. 5.13: Amplitude in response to a stimulus of strength a and frequency ω for $\mu = 0.2$

input amplitudes a. It can be seen that when a is small, there is significant amplification and sharp frequency tuning around the resonant frequency ω_0. The amplitude response curves are qualitatively similar to those seen experimentally [287, 423]. At the resonant frequency ($\omega = \omega_0$), the amplitude equation reduces to the simpler form

$$A^3 - \mu A = a,$$

which establishes that at the Hopf bifurcation point where $\mu = 0$, we have

$$A \sim a^{1/3}.$$

This is a highly compressive nonlinearity that boosts weak signals much more strongly than strong signals. On the other hand, if the stimulus frequency differs significantly from ω_0, then the cubic term in the amplitude equation can be neglected, and the system operates in a linear regime for which

$$A \sim \frac{a}{\omega - \omega_0},$$

and there is a 90° phase lag. Finally, note that in order for the above nonlinear resonance to be realized by hair bundles, there has to be some feedback mechanism that keeps the system close to Hopf bifurcation point for a range of different natural frequencies—a process known as *self-tuning* [90, 158].

5.5 Exercises

Problem 5.1 (Physical limits of biochemical signaling). Consider a receptor at position $\mathbf{x}_0$ with a single binding site. Let $n(t)$ denote the fractional occupancy of the site and $c(x,t)$ the concentration of diffusing ligand. Binding and diffusion are coupled according to equations (5.2.10) and (5.2.11).

(a) By linearizing these equations about the uniform steady state $(\bar{n}, \bar{c})$ and using detailed balance, show that

$$\gamma \frac{d\delta n}{dt} = -\kappa \delta n + k_+[1-\bar{n}]\gamma \delta c(\mathbf{x}_0, t) + \delta F,$$

with

$$\gamma = \frac{k_B T}{k_+ \bar{c}(1-\bar{n})}, \quad \kappa = (k_+ \bar{c} + k_-)\gamma,$$

and

$$\frac{\partial \delta c(\mathbf{x},t)}{\partial t} = D\nabla^2 \delta c(\mathbf{x},t) - \delta(\mathbf{x}-\mathbf{x}_0)\frac{d\delta n(t)}{dt}.$$

(b) By Fourier transforming the diffusion equation with respect to space and time, with

$$\delta c(\mathbf{k},\omega)=\int_{-\infty}^{\infty}dt\int d^3x\,\mathrm{e}^{i\omega t}\,\mathrm{e}^{i\mathbf{k}\cdot x}\delta c(\mathbf{x},t),\quad \delta n(\omega)=\int_{-\infty}^{\infty}\mathrm{e}^{i\omega t}\delta n(t)dt$$

show that

$$\delta c(\mathbf{x},\omega)=i\omega\Sigma_0(\omega)\delta n(\omega),\quad \Sigma_0(\omega)=\int\frac{\mathrm{e}^{i\mathbf{k}\cdot(\mathbf{x}_0-\mathbf{x})}}{Dk^2-i\omega}\frac{d^3k}{(2\pi)^3}.$$

(*c*) Hence, derive the result

$$\frac{\delta n(\omega)}{\delta F(\omega)}=\frac{1}{\gamma}\frac{1}{-i\omega[1+\Sigma(\omega)]+k_+\bar{c}+k_-},$$

where $\Sigma(\omega)=k_+(1-\bar{n})\Sigma_0(\omega)$.

Problem 5.2 (Ising model of receptor clustering.). Consider the Ising model of receptor clustering with partition function

$$Z=\sum_{\mathbf{a}}\mathrm{e}^{-H(\mathbf{a})/k_BT},\quad H(\mathbf{a})=-J\sum_{\langle m,n\rangle}(2a_m-1)(2a_n-1)+F([L])\sum_m a_m.$$

Using the identity

$$\left\langle\sum_m a_m\right\rangle=-\frac{\partial\ln Z}{\partial F},$$

and the analysis of the Ising model in Sect. 4.5, show that the mean level of kinase activity per receptor is

$$\langle a\rangle=\frac{1}{2}\left[1-\frac{\sinh F([L])}{\sqrt{\sinh^2F([L])+\mathrm{e}^{-4J}}}\right],$$

with

$$F([L])=\Delta E+k_BT\ln\frac{1+K_i[L]}{1+K_a[L]}.$$

Given that $K_i>K_a$, describe how $\langle a\rangle$ changes as $[L]$ increases from 0 to ∞.

Problem 5.3 (Linear response in the Barkai–Leibler model). Consider the Barkai–Leibler model of adaptation in bacterial chemotaxis. The methylation level evolves as

$$\frac{dv}{dt}=F(a)\equiv\Gamma_R-\frac{\Gamma_Ba}{K_B+a},$$

where $\Gamma_{R,B}$ are maximum catalytic rates and the average receptor activity a is given by the MWC model (with $k_BT = 1$):

$$a = a([L], \nu) \equiv \left[1 + \exp\left(N\left(\Delta E(\nu) + \ln\frac{1 + K_i[L]}{1 + K_a[L]}\right)\right)\right]^{-1}.$$

Let a_0, ν_0 be the steady state at a background ligand concentration L_0. Suppose that there is a small oscillatory modulation of the ligand concentration

$$[L](t) = L_0 e^{A\cos(\omega t)},$$

where A is the amplitude, $A \ll 1$, and ω is the modulation frequency. Finally, assume that $K_a[L] \ll 1 \ll K_i[L]$ so that

$$\ln\frac{1 + K_i[L]}{1 + K_a[L]} \approx \ln([L]/K_i).$$

(a) Linearizing about the steady state (a_0, ν_0), show that

$$\frac{d\Delta\nu}{dt} = F'(a_0)\Delta a,$$

with $\Delta\nu = \nu - \nu_0$,

$$\Delta a = a - a_0 = Na_0(1 - a_0)[\alpha\Delta\nu - A\cos(\omega t)],$$

and $\alpha = \Delta E'(\nu_0)$.

(b) Setting

$$\Delta\nu = \mathrm{Re}[A_m e^{i\omega t}], \quad \Delta a = \mathrm{Re}[A_a e^{i\omega t}],$$

use part (a) to solve for the complex amplitudes A_m, A_a:

$$A_a = \frac{i\omega c_a}{i\omega + \omega_m}A, \quad A_m = \frac{\omega_m c_m}{i\omega + \omega_m}A,$$

where

$$c_a = Na_0(1 - a_0), \quad c_m = \alpha^{-1}, \quad \omega_m = -\alpha F'(a_0)Na_0(1 - a_0).$$

(*c*) The linear response of receptor activity can be characterized by the amplitude $|A_a|$ and phase $\phi_a = \pi/2 + \tan^{-1}(\nu/\nu_m)$. Plot $|A_a|/|A_a|_{\max}$ and ϕ/π as a function of ν/ν_m. (A typical value of ν_m is around 5×10^{-3} Hz.)

Problem 5.4 (Telegraph equation). Consider the simple velocity-jump process

$$\frac{\partial p_+}{\partial t} + v\frac{\partial p_+}{\partial x} = -\lambda p_+ + \lambda p_-,$$
$$\frac{\partial p_-}{\partial t} - v\frac{\partial p_-}{\partial x} = \lambda p_+ - \lambda p_-,$$

for constant v, λ.

(a) By adding and subtracting the pair of equations, derive corresponding equations for the total probability density $p = p_+ + p_-$ and the flux $j = v[p_- - p_-]$. Eliminate j by cross-differentiating in order to derive the telegraph equation

$$\frac{\partial^2 p}{\partial t^2} + 2\lambda \frac{\partial p}{\partial t} = v^2 \frac{\partial^2 p}{\partial x^2}.$$

(b) By performing the change of variables $u(x,t) = p(x,t)e^{\lambda t}$, show that u satisfies another version of the telegraph equation

$$\frac{\partial^2 u}{\partial t^2} - \lambda^2 u = v^2 \frac{\partial^2 u}{\partial x^2}.$$

Problem 5.5 (Moment equations for a 1D model of bacterial chemotaxis). Consider the 1D model of chemotaxis given by equations (5.3.18).

(a) Show how multiplying equations (5.3.18) by 1 or z, integrating with respect to z, and adding or subtracting the resulting equations leads to the hierarchy of moment equations (5.3.20) for the macroscopic variables defined by equations (5.3.19).

(b) In parts (b)–(d) set $j_2 = 0$ and $c'(x) = c_0$ in equations (5.3.20). Taking first and second moments of equation (5.3.20a) with respect to x, derive equation (5.3.22).

(c) Taking the zeroth moments of equations (5.3.20b–d), derive a system of first-order equations for the triplet (j_0, n_{10}, j_{10}) and show that there exists a unique stable fixed point for which

$$j_0 = \frac{\beta v^2 c_0 \tau_2}{\lambda_0 + 2\lambda_0^2 \tau_2}.$$

(d) Similarly, taking the first moment of equations (5.3.20b–d), derive a system of first-order equations for the triplet (j_1, n_{11}, j_{11}) and show that there exists a unique stable fixed point for which

$$2j_1 - 2\bar{x}j_0 = \left(\frac{v^2}{\lambda_0} + \frac{2\beta v^4 c_0^2 \tau_2^3}{(\lambda_0 + 2\lambda_0^2 \tau_2)^2} \right).$$

(e) Use parts (b)–(d) to obtain the asymptotic drift and diffusion coefficient of equation (5.3.24).

Problem 5.6 (Dynamical model of hair cell oscillations). Consider the stochastic version of the planar model of hair cell oscillations given by equations (5.4.5) and (5.4.7) in the QSS limit of fast Ca^{2+} dynamics:

$$\xi \frac{dx}{dt} = -N\left[k_{\rm sp}(x - x_M - \delta P_0(x)/\gamma) + k_{\rm sp}x\right] + \eta(t)$$

and

$$\xi_M \frac{dx_M}{dt} = N k_{\rm gs}\left[x - x_M - \delta P_0(x)/\gamma\right] - f_{\max}(1 - SP_o(x)) + \eta_M(t),$$

with the open channel probability $P_0(x)$ given by equation (5.4.6) and Gaussian white noise terms satisfying

$$\langle \eta(t)\eta(0)\rangle = 2k_BT\xi\,\delta(t)$$
$$\langle \eta_M(t)\eta_M(0)\rangle = 2k_BT_M\xi_M\,\delta(t)$$

for $T_M = 1.5T$.

(a) First consider the zero-noise case. Investigate the existence and stability of any fixed points as a function of $f_{\max}$ for the parameter values given in Fig. 5.11 in the two cases (a) $S = 0.5$ and (b) $S = 1.0$. Show that the results are consistent with the bifurcation diagram shown in Fig. 5.11.
(b) Using the numerical methods outlined in Sect. 2.6.6, simulate the stochastic version of the model for $f_{\max} = 50.3pn$ and $S = 0.65$, and plot the power spectrum.

Problem 5.7 (Biochemical amplification in photoreceptors.). One of the remarkable features of a photoreceptor in the retina is that it can detect individual photons (particles) of light. This is mediated by a single rhodopsin molecule that is activated by the photon and generates a current via a signaling cascade. Since the lifetime of the active state of rhodopsin is stochastic, how does the molecule produce a reliable single-photon response? One possible mechanism for enhanced reliability is based on the observation that rhodopsin is inactivated in a series of phosphorylation steps [322].

(a) Suppose that a rhodopsin molecule is activated at time $t = 0$ and that the rate of decay to the inactive state is k. The probability $P_0(t)$ that it is still active at time t satisfies the equation

$$\frac{dP_0}{dt} = -kP_0,$$

with $P_0(0) = 1$. The probability density that the molecule is active for exactly a time t is then $kP_0(t)$. Let the mean and variance of the activation time be μ and σ, respectively. Show that $\sigma/\mu = 1$.
(b) Now suppose that upon light activation, rhodopsin starts in the state R_0 and undergoes a sequence of phosphorylation steps according to the reaction scheme

$$R_0 \overset{k_0}{\to} R_1 \overset{k_1}{\to} R_2 \cdots \overset{k_{n-2}}{\to} R_{n-1} \overset{k_{n-1}}{\to} \text{inactive rhodopsin},$$

where R_j is the state with j sites phosphorylated. The rhodopsin is inactivated as soon as a maximum of n sites are phosphorylated. Write down the corresponding set of kinetic equations for the probability $P_k(t)$, $k = 0,\ldots,n-1$, that the molecule is in state k at time t.
(c) Assume that each state R_j elicits the same response. The relevant quantity is then the probability density that the molecule exits the state R_{n-1} exactly at time t, which is given by $kP_{n-1}(t)$. By Fourier transforming the kinetic equations with respect to time t, show that

$$P_{n-1}(t) = \int_{-\infty}^{\infty} \frac{k^{n-1}}{(k+i\omega)^n} \mathrm{e}^{i\omega t} \frac{d\omega}{2\pi}.$$

Using the identity

$$\int_0^{\infty} t^{n-1} \mathrm{e}^{-kt} \mathrm{e}^{-i\omega t} dt = \frac{(n-1)!}{(k+i\omega)^n}.$$

determine $P_{n-1}(t)$, and then calculate

$$\mu = k \int_0^{\infty} P_{n-1}(t) t dt, \quad \sigma^2 = k \int_0^{\infty} P_{n-1}(t) t^2 dt - \mu^2,$$

Hence, deduce that

$$\frac{\sigma}{\mu} = \frac{1}{\sqrt{n}}.$$

This establishes that increasing the number of phosphorylation steps n increases the reliability of the single photon response.

Chapter 6
Stochastic Gene Expression and Regulatory Networks

Genetically identical cells exposed to the same environmental conditions can show significant variation in molecular content and marked differences in phenotypic characteristics. This intrinsic variability is linked to the fact that many cellular events at the genetic level involve small numbers of molecules (low copy numbers). We have already encountered intrinsic noise effects within the context of stochastic ion channels (Chap. 3) and biochemical signaling (Chap. 5). Although stochastic gene expression was originally viewed as having detrimental effects on cellular function, with potential implications for disease, it is now seen as being potentially advantageous. For example, intrinsic noise can provide the flexibility needed by cells to adapt to fluctuating environments or respond to sudden stresses and can also support a mechanism by which population heterogeneity is established during cell differentiation and development. Since the demonstration of a functional role for stochastic gene expression in λ-phage [13], there has been an explosion of studies focused on investigating the origins and consequences of noise in gene expression (see the reviews [312, 408, 502, 521, 555, 644]). This typically involves establishing the molecular mechanisms of noise generation at the single gene level and then building on this knowledge to test and predict its effects on larger regulatory networks. *Gene regulation* refers to the cellular processes that control the expression of proteins, dictating under what conditions specific proteins should be produced from their parent DNA. This is particularly crucial for multicellular organisms, where all cells share the same genomic DNA, yet do not all express the same proteins. That is, selective gene expression allows the cells to specialize into different phenotypes (cell differentiation), resulting in the development of different tissues and organs with distinct functional roles.

In this chapter we explore the effects of noise on gene expression and protein synthesis. We begin by reviewing the basic steps in gene expression (Sect. 6.1). We then analyze transcription and translation in some simple unregulated networks and show how translational bursts in the production of protein can occur (Sect. 6.2). Various simple gene regulatory networks are analyzed in Sect. 6.3 using the linear noise (diffusion) approximation (see also Sect. 3.2) and Fourier (spectral) methods. Important examples of nonlinear feedback regulatory networks such as genetic

P.C. Bressloff, *Stochastic Processes in Cell Biology*, Interdisciplinary Applied Mathematics 41, DOI 10.1007/978-3-319-08488-6_6

switches and genetic oscillators are studied in Sect. 6.4, including the *lac* operon and the genetic circuits of the circadian clock. We also discuss some methods for analyzing the effects of noise on biochemical oscillators. The efficacy of gene networks in transmitting information in the presence of molecular noise is investigated in Sect. 6.5, where some basic concepts such as Shannon information and mutual information are introduced. We then look at some models of kinetic proofreading, which is a mechanism for increasing the fidelity of molecular recognition during protein synthesis, for example, and other cellular processes such as T-cell activation in immunology (see Sect. 6.7). Finally, the stochastic simulation algorithm (SSA) introduced by Gillespie to simulate sample trajectories of a gene or biochemical network is described in Sect. 6.8.

6.1 Basics of Gene Expression

In Fig. 6.1a we show the two main stages in the expression of a single gene according to the *central dogma*.

1. *Transcription (DNA → RNA).* The first major stage of gene expression is the synthesis of a *messenger RNA* (mRNA) molecule with a nucleotide sequence complementary to the DNA strand from which it is copied—this serves as the template for protein synthesis. Transcription is mediated by a molecular machine known as RNA polymerase (RNAP). In the case of eukaryotes, transcription takes place in the cell nucleus, whereas subsequent protein synthesis takes place in the cytoplasm, which means that the mRNA has to be exported from the nucleus as an intermediate step.
2. *Translation (RNA → protein).* The second major stage is synthesis of a protein from mRNA. Translation is mediated by a macromolecule known as a *ribosome*, which produces a string of amino acids (polypeptide chains), each specified by a *codon* (represented by three letters) on the mRNA molecule. Since there are four nucleotides (A, U, C, G), there are 64 distinct codons, e.g., AUG and CGG, most of which code for a single amino acid. The process of translation consists of ribosomes moving along the mRNA without backtracking (from one end to the other, technically known as the 5' end to the 3' end) and is conceptually divided into three major stages (as is transcription): initiation, elongation, and termination. Each elongation step invokes translating or "reading" of a codon and the binding of a freely diffusing transfer RNA (tRNA) molecule that carries the specific amino acid corresponding to that codon. Once the chain of amino acids has been generated a number of further processes occur in order to generate a correctly folded protein.

The above simplified picture ignores a major feature of cellular processing, namely, gene regulation. Individual cells frequently have to make "decisions," that is, to express different genes at different spatial locations and times and at different activity levels. One of the most important mechanisms of genetic control is transcriptional regulation, that is, determining whether or not an mRNA molecule

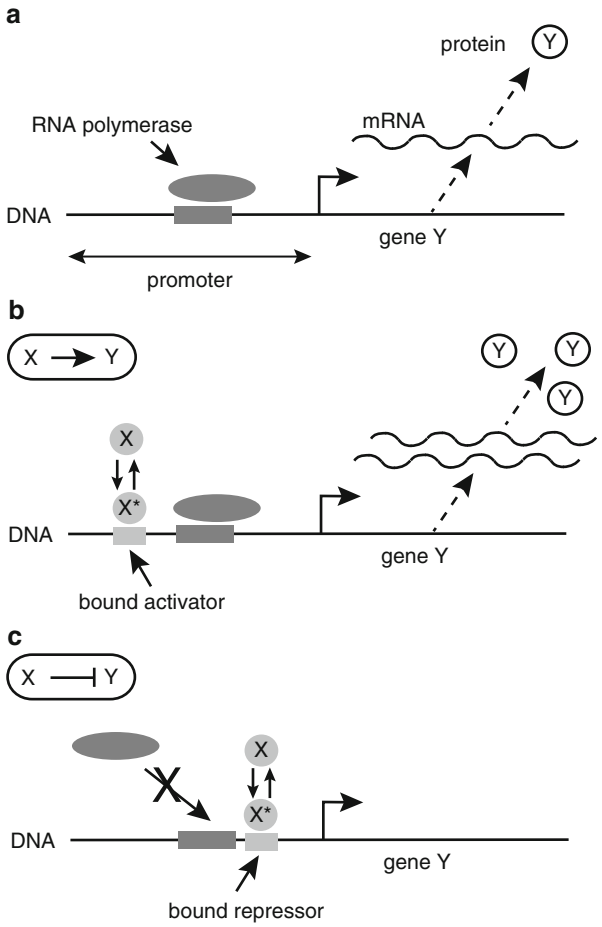


Fig. 6.1: Transcriptional regulation due to the binding of a repressor or activator protein to a promoter region along the DNA. (**a**) Unregulated transcription of a gene Y following binding of RNA polymerase to the promoter region. The resulting mRNA exits the nucleus and is then translated by ribosomes to form protein Y. (**b**) Increased transcription due to the binding of an activator protein *X* to the promoter. An activator typically transitions between inactive and active forms; the active form X^* has a high affinity to the promoter binding site. An external chemical signal can regulate transitions between the active and inactive states. (**c**) Transcription can be stopped by a repressor protein X binding to the promoter and blocking the binding of RNA polymerase

is made. The control of transcription (switching on or off a gene) is mediated by proteins known as transcription factors (see Fig. 6.1b, c). Negative control (or repression) is mediated by repressors that bind to a promoter region along the DNA where RNAP has to bind in order to initiate transcription—it thus inhibits transcription. On the other hand, positive control (activation) is mediated by activators that increase the probability of RNAP binding to the promoter. The presence of

transcription factors means that cellular processes can be controlled by extremely complex gene networks, in which the expression of one gene produces a repressor or activator, which then regulates the expression of the same gene or another gene. This can result in many negative and positive feedback loops, the understanding of which lies at the heart of systems biology [5]. In addition to transcriptional regulation, there are a variety of other mechanisms that can control gene expression including mRNA and protein degradation and translational regulation.

6.1.1 Intrinsic Versus Extrinsic Noise Sources

Following Swain et al. [164], it is useful to distinguish between contributions arising from fluctuations that are inherent to a given system of interest (intrinsic noise) from those arising from external factors (extrinsic noise). In the model of gene expression shown in Fig. 6.1, intrinsic noise is due to fluctuations generated by the binding/unbinding of a repressor or activator and mRNA and protein production and decay—these can be significant due to the small number of molecules involved. Extrinsic noise sources are defined as fluctuations and population variability in the rate constants associated with these events. The classification of a noise source as intrinsic rather than extrinsic is context-dependent, so that intrinsic noise at one level can act as extrinsic noise at another level. Gene-intrinsic noise refers to the variability generated by molecular-level noise in the reaction steps that are intrinsic to the process of gene expression. Network-intrinsic noise is generated by fluctuations and variability in signal transduction and includes gene-intrinsic noise in the expression of regulatory genes. Cell-intrinsic noise arises from gene-intrinsic noise and network-intrinsic noise, as well as fluctuations and variability in cell-specific factors, such as the activity of ribosomes and polymerases, metabolite concentrations, cell size, cell age, and stage of the cell cycle.

An operational definition of gene-intrinsic noise is the difference in the expression of two almost identical genes from identical promoters in single cells averaged over a large cell population. This definition is based on the assumptions that the two genes are affected identically by fluctuations in cell-specific factors and that their expression is perfectly correlated if these fluctuations are the only source of population heterogeneity. The contribution of gene-intrinsic noise can then be investigated experimentally using two-reporter assays (see Sect. 1.2). These assays evaluate, in single cells, the difference in the abundances of two equivalent reporters, such as red and green fluorescent protein, expressed from identical promoters, located at equivalent chromosomal positions. This allows measurements of noise fluctuations generated by the biochemical reaction steps that are intrinsic to the process of gene expression, and how this is affected by mutations or gene deletions. There are, however, some potential limitations. For example, contributions from extrinsic factors, such as imperfect timing in replication and intracellular heterogeneity, might

be measured as gene-intrinsic noise. Moreover, because increased variability in regulatory signals might cause cells to adapt distinct expression states, the measured population-average gene-intrinsic noise and the extrinsic regulatory noise might not always be independent.

6.1.2 Biological Significance of Stochasticity

Stochasticity in gene expression is generally believed to be detrimental to cell function, because fluctuations in protein levels can corrupt the quality of intracellular signals, negatively affecting cellular regulation. One possible benefit of randomness, however, is that it can provide a mechanism for phenotypic and cell-type diversification:

1. Stochasticity in gene expression that generates phenotypic heterogeneity is expected to be particularly beneficial to microbial cells that need to adapt efficiently to sudden changes in environmental conditions. Fluctuations in gene expression provide a mechanism for 'sampling' distinct physiological states and could therefore increase the probability of survival during times of stress, without the need for genetic mutation. A classical example is the infection of *E. coli* by a bacterial virus known as *lambda phage*. Infection is governed by a particular lysis/lysogeny decision circuit, in which only a fraction of infecting phage chooses to lyse (break down) the cell. The remainder become dormant lysogens, in which the bacteriophage nucleic acid is integrated into the host bacterium's genome, awaiting bacterial stress signals to enter the production phase of their life cycle.
2. Switching between phenotypic states with different growth rates might be an important factor in the phenomenon of persistent bacterial infections after treatment with antibiotics. Although most of the population is rapidly killed by the treatment, a small genetically identical subset of dormant 'persistor' cells can survive an extended period of exposure. When the drug treatment is removed, the surviving persistors randomly transition out of the dormant state, causing the infection to reemerge.
3. The primary purpose of the *Saccharomyces cerevisiae* (yeast) galactose-utilization network is to increase the uptake and metabolism of galactose. It involves several positive feedback loops that generate bistability in the network, which endow cells (and their progeny) with long-term epigenetic memory of past galactose-consumption states. It has been suggested that the existence of a negative feedback loop (which appear spurious from a deterministic perspective) reduces this memory by increasing the rate at which cells randomly switch between different phenotypic states that are associated with different expression of the galactose-utilization genes. As a result, the biological function of negative feedback might be to prevent cells from being trapped in suboptimal phenotypic states.

4. Stochasticity may also play a constructive role in development and cellular differentiation in higher organisms. For example, during *Drosophila melanogaster* development, stochastic fluctuations in the turnover of two proteins, Notch and Delta, might underlie the random emergence of neural precursor cells from an initial homogeneous cell population.

6.2 Unregulated Transcription and Translation

The key steps in transcription are binding of RNAP (P) to the relevant promoter region of DNA (D) to form a closed complex (PD_c), the unzipping of the two strands of DNA to form an open complex (PD_o), and finally promoter escape, when RNAP reads one of the exposed strands

$$P + D \underset{k_-}{\overset{k_+}{\rightleftharpoons}} PD_c \xrightarrow{k_{\text{open}}} PD_0 \xrightarrow{k_{\text{escape}}} \text{transcription}.$$

Once the RNAP is reading the strand, the promoter is unoccupied and ready to accept a new polymerase. The binding/unbinding of polymerase is very fast, $k_\pm \gg k_{\text{open}}$ so that the first step happens many times before formation of an open complex. Hence, one can treat the RNAP as in quasi-equilibrium with the promoter characterized by an equilibrium constant $K_P = k_+/k_-$. The rate of transcription will thus be proportional to the fraction of bound RNAP, $k_+/(k_+ + k_-)$. The production of mRNA from a typical gene in *E. coli* occurs at a rate around 10 per minute, while the average lifetime of mRNA due to degradation is around a minute. This implies that on average there are ten mRNA molecules per cell. Generation of the mRNA molecule occurs at a rate of 50 nucleotides per second. Hence, a typical gene of around 1,000 nucleotides will be transcribed in about 20 s. Thus, there are around three RNAP per gene at any one time, suggesting the number fluctuations will be significant.

First, suppose that we ignore any regulation of the promoter as in Fig. 6.1a, and collapse the various stages of transcription into a single step with mRNA production rate k. Letting γ denote the rate of mRNA degradation and $m(t)$ the number of mRNA molecules at time t, we have the reaction

$$m \xrightarrow{k} m+1, \quad m \xrightarrow{\gamma} m-1$$

with corresponding kinetic equation for the concentration $x = m/\Omega$, where Ω is cell volume

$$\frac{dx}{dt} = k - \gamma x.$$

Clearly, given that m is of order 10, the law of mass action breaks down and we have to consider the corresponding birth–death master equation

$$\frac{dp_m(t)}{dt} = -\Omega k p_m(t) + \Omega k p_{m-1}(t) - \gamma m p_m(t) + \gamma(m+1)p_{m+1}(t) \tag{6.2.1}$$

for $m \geq 0$ and $P_{-1} \equiv 0$. This is identical to the autonomous version (3.6.3) of the master equation for a stochastic gating model (Sect. 3.6). We immediately deduce that the resulting probability density is given by the Poisson distribution (3.6.7). Hence, in the limit $t \to \infty$ we obtain a stationary Poisson process with

$$p_m = \mathrm{e}^{-\lambda}\frac{\lambda^m}{m!}, \quad \lambda = \Omega k/\gamma. \tag{6.2.2}$$

It follows that

$$\langle m \rangle = \lambda, \quad \mathrm{var}[m] = \lambda.$$

This is an important result because both the mean and variance in the number of mRNA molecules can be measured experimentally. One commonly used measure of the level of noise in a regulatory network is the so-called Fano factor:

$$\text{Fano factor} = \frac{\langle m^2 \rangle - \langle m \rangle^2}{\langle m \rangle}. \tag{6.2.3}$$

For the unregulated process, the Fano factor is one.

6.2.1 Translational Bursting

In addition to transcription, other steps in the central dogma are also subject to variability including protein translation, which often occurs in bursts [36, 89, 199, 427]. One could add the translation step (mRNA $\to$ protein) to the previous model. However, it is simpler to proceed by exploiting the fact that a single mRNA molecule has a much shorter lifetime than a protein. First, consider a single mRNA molecule with a degradation rate γ, which starts synthesizing a protein at time $t = 0$. Let $p_0(n,t)$ ($p_c(n,t)$) denote the probability that there are n proteins at time t and the mRNA has not (has) decayed. Neglecting protein degradation, we have the master equation

$$\frac{dp_0(n,t)}{dt} = -\gamma p_0(n,t) + r[p_0(n-1,t) - p_0(n,t)] \tag{6.2.4a}$$

$$\frac{dp_c(n,t)}{dt} = \gamma p_0(n,t), \tag{6.2.4b}$$

where r is the rate of protein production and $p_0(-1,t) \equiv 0$. Let

$$P(n) = \lim_{t\to\infty} p_c(n,t).$$

Note that $\lim_{t\to\infty} p_0(n,t) = 0$ due to the decay of mRNA. Integrating Eq. (6.2.4b) with respect to time gives

$$P(n) = \gamma \int_0^\infty p_0(n,t)dt,$$

since $p_c(n,0) = 0$. In order to compute $p_0(n,t)$, integrate Eq. (6.2.4a) with respect to time using $p_0(n,t) = \delta_{n,0}$:

$$-\delta_{n,0} = -P(n) + \frac{r}{\gamma}[P(n-1) - P(n)].$$

Setting $n = 0$ gives

$$P(0) = \frac{\gamma}{r+\gamma}.$$

For $n \geq 1$, we have the recurrence relation

$$P(n) = \frac{r}{r+\gamma}P(n-1) \implies P(n) = \left(\frac{r}{r+\gamma}\right)^n \frac{\gamma}{r+\gamma}.$$

An important quantity is the so-called burst size b, which is the mean number of proteins produced per mRNA. Using generating functions it can be shown that (see Ex. 6.1)

$$b = \frac{r}{\gamma}.$$

The idea of a translational burst refers to the observation that a single mRNA generates a burst of protein production before it decays (see Fig. 6.2a).

Now suppose that there are m mRNA molecules and that translation of each mRNA proceeds independently. The probability of producing N proteins due to bursts from each mRNA molecule can be expressed as a multiple convolution [509]. For example, if $m = 2$, then

$$P_2(N) = \sum_{n=0}^{N} P(n)P(N-n),$$

and

$$P_3(N) = \sum_{n=0}^{N} P(n) \sum_{n'=0}^{N-n} P(n')P(N-n-n').$$

Assume that the number of proteins is sufficiently large so that we can approximate the sums by integrals, for example,

$$P_2(N) = \int_0^N P(n)P(N-n)dn.$$

The advantage of the integral formulation is that one can use Laplace transforms and the convolution theorem. Thus, setting

$$\tilde{P}_m(s) = \int_0^\infty P_m(n)\mathrm{e}^{-sn}dn,$$

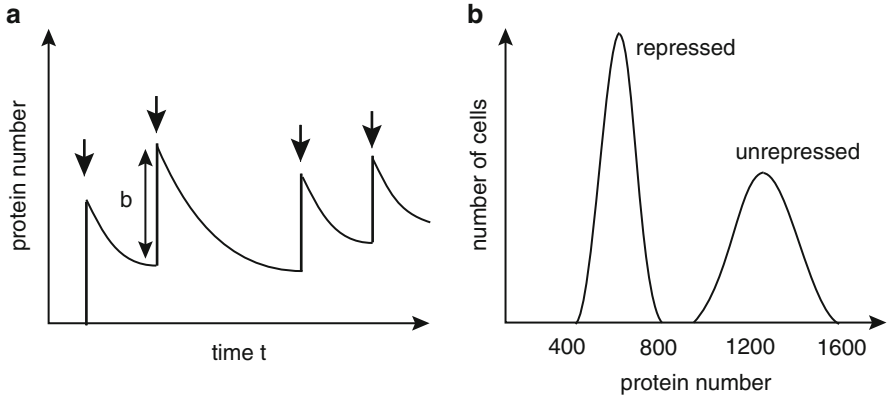


Fig. 6.2: Effects of noise in gene expression. (**a**) Schematic illustration of translational bursting. Each *arrow* represents a burst event where an mRNA transcript releases a burst of proteins of average size b, and proteins decay between bursts. (**b**) Illustration of how negative feedback in an autoregulatory network reduces the mean number of proteins but also reduces the size of fluctuations

we have

$$\tilde{P}_m(s) = \left[\tilde{P}(s)\right]^m.$$

Calculating $\tilde{P}_m(s)$ and then inverting yields the result (see Ex. 6.1)

$$P_m(n) = \left(\frac{b}{1+b}\right)^n \left(\frac{1}{1+b}\right)^m \frac{n^{m-1}}{\Gamma(m)}.$$

For $n, b \gg 1$, we can make the approximation

$$\left(\frac{b}{1+b}\right)^n = \mathrm{e}^{-n\ln(1+b^{-1})} \approx \mathrm{e}^{-n/b},$$

which leads to the gamma distribution for n with m fixed:

$$P_m(n) \equiv F(n; m, b^{-1}) = \frac{n^{m-1}\mathrm{e}^{-n/b}}{b^m \Gamma(m)}. \tag{6.2.5}$$

From properties of the gamma distribution, we immediately note that for a given number of mRNA molecules,

$$\langle n \rangle = mb, \quad \mathrm{var}(n) = mb^2.$$

Hence, under the various approximations the Fano factor is of the order of the burst size b. Finally, an estimate for m is $m \approx k/\gamma_0$ where k is the rate of production of mRNAs and γ_0 is the frequency of the cell cycle (assuming that it is higher than the rate of protein degradation).

An alternative approach to analyzing protein bursting is to start from the Chapman–Kolmogorov (CK) equation [199]

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial}{\partial x}[\gamma_0 x p(x)] + k\int_0^x w(x-x')p(x',t)dx', \tag{6.2.6}$$

where $p(x,t)$ is the probability density for x protein molecules (treating x as a continuous variable) at time t, and

$$w(x) = \frac{1}{b}\mathrm{e}^{-x/b} - \delta(x). \tag{6.2.7}$$

The first term on the right-hand side of the CK equation represents protein degradation, where the second term represents the production of proteins from exponentially distributed bursts. The gamma distribution (6.2.5) with $n \to x$ is obtained as the stationary solution of the CK equation, which can be established using Laplace transforms (Ex. 6.2). It is also possible to incorporate autoregulatory feedback into the CK equation by allowing the burst rate to depend on the current level of protein x, which acts as its own transcription factor [199]:

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial}{\partial x}[\gamma_0 x p(x)] + k\int_0^x w(x-x')c(x')p(x',t)dx'. \tag{6.2.8}$$

One possible form of the response function $c(x)$ is a Hill function

$$c(x) = \frac{k^s}{k^s + x^s},$$

with $s > 0$ ($s < 0$) corresponding to negative (positive) feedback. In this case, the stationary density takes the form (Ex. 6.2)

$$p(x) = Ax^{m(1+\varepsilon)-1}\mathrm{e}^{-x/b}[1 + (x/k)^s]^{-m/s}.$$

A more general mathematical analysis of bursting in discrete and continuous models can be found in [407].

6.3 Simple Models of Gene Regulation

One of the simplest gene regulatory networks consists of a gene that can be in one of two states, active or inactive (see Fig. 6.3). In the active state the gene produces protein X at a rate r, which subsequently degrades at a rate γ, whereas no protein is produced in the inactive state. For simplicity, the stages of transcription and translation are lumped together so we do not keep track of the amount of mRNA. Moreover, the transcription factor Y that switches on the gene is independent of protein X, that is, there is no feedback. Since the rate of activation k_+ will be proportional to the concentration c of Y in the nucleus, this simple network can be viewed as an

input/output device that converts the input signal c to an output signal given by the concentration x of protein X. Moreover, if X is a green fluorescent protein, then the output response can be measured. In Sect. 6.5, we will consider how effective such a feedforward networks is in transmitting information in the presence of molecular noise, following the work of Tkacik et al. [631, 632, 634, 665]. Here we will focus on calculating the level of noise.

6.3.1 Transcriptional Bursting in a Two-State Model

The reaction scheme of the regulatory network shown in Fig. 6.3 is

$$I \underset{k_-}{\overset{k_+}{\rightleftharpoons}} A \xrightarrow{r} \mathrm{p} \xrightarrow{\gamma} \emptyset,$$

where A and I denote the active and inactive states of the gene. We first consider the case in which the number of X proteins is sufficiently large so that we can represent the dynamics in terms of a continuous-valued protein concentration x [318]. The latter evolves according to the (piecewise) deterministic equation

$$\frac{dx}{dt} = rn(t) - \gamma x, \tag{6.3.1}$$

where the discrete random variable $n(t)$ represents the current state of the gene with $n(t) = 1$ (active) or $n(t) = 0$ (inactive). We thus have another example of a stochastic hybrid system. Let $p_j(x,t)$ denote the probability density of the protein concentration for $n(t) = j$, $j = 0, 1$. We then have the differential Chapman–Kolmogorov (CK) equation

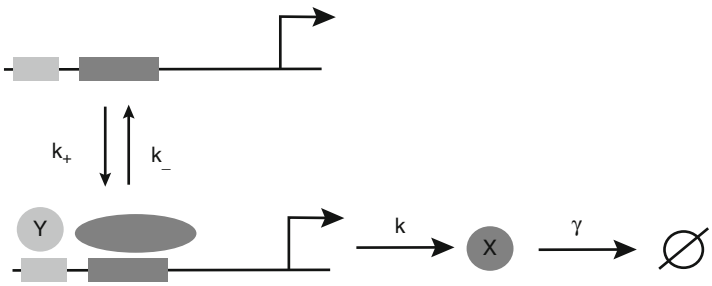


Fig. 6.3: Simple example of a two-state gene regulatory network. The promoter transitions between an active state (bound by a transcription factor protein Y and RNA polymerase) and an inactive state with rates $k_\pm$. The active state produces protein X at a rate r and protein X degrades at a rate γ

$$\frac{\partial p_0}{\partial t} = -\frac{\partial}{\partial x}(-\gamma x p_0(x,t)) + k_- p_1(x,t) - k_+ p_0(x,t) \tag{6.3.2a}$$

$$\frac{\partial p_1}{\partial t} = -\frac{\partial}{\partial x}([r-\gamma x]p_1(x,t)) + k_+ p_0(x,t) - k_- p_1(x,t), \tag{6.3.2b}$$

supplemented by the no-flux boundary conditions $J_s(x) = 0$ at $x = 0, r/\gamma$, where $J_0(x) = -\gamma x p_0(x)$ and $J_1(x) = [r-\gamma x]p_1(x)$. In the limit that the switching between active and inactive states is much faster than the protein dynamics, the probability that the gene is active rapidly converges to the steady state $k_+/(k_+ + k_-)$, and we obtain the deterministic equation

$$\frac{dx}{dt} = r\langle n \rangle - \gamma x = \frac{rk_+}{k_+ + k_-} - \gamma x. \tag{6.3.3}$$

Following [318], we will characterize the long-time behavior of the system in terms of the steady-state solution, which satisfies

$$\frac{d}{dx}(-\gamma x p_0(x)) = k_- p_1(x) - k_+ p_0(x) \tag{6.3.4a}$$

$$\frac{d}{dx}([r-\gamma x]p_1(x)) = k_+ p_0(x) - k_- p_1(x). \tag{6.3.4b}$$

The no-flux boundary conditions imply that $p_0(r/\gamma) = 0$ and $p_1(0) = 0$. First, note that we can take $x \in [0, r/\gamma]$ and impose the normalization condition

$$\int_0^{r/\gamma} [p_0(x) + p_1(x)]dx = 1.$$

Integrating Eq. (6.3.4) with respect to x then leads to the constraints

$$\int_0^{r/\gamma} p_0(x)dx = \frac{k_-}{k_- + k_+}, \quad \int_0^{r/\gamma} p_1(x)dx = \frac{k_+}{k_- + k_+}.$$

Adding Eqs. (6.3.4a) and (6.3.4b) we can solve for $p_0(x)$ in terms of $p_1(x)$ and then generate a closed differential equation for $p_1(x)$. We thus obtain a solution of the form (see Ex. 6.3),

$$p_0(x) = C(\gamma x)^{-1+k_+/\gamma}(r-\gamma x)^{k_-/\gamma}, \quad p_1(x) = C(\gamma x)^{k_+/\gamma}(r-\gamma x)^{-1+k_-/\gamma} \tag{6.3.5}$$

for some constant C. Imposing the normalization conditions then determines C as

$$C = \gamma \left[r^{(k_+ + k_-)/\gamma} B(k_+/\gamma, k_-/\gamma) \right]^{-1},$$

where $B(\alpha, \beta)$ is the beta function:

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt.$$

Finally, setting $r/\gamma = 1$, the total probability density $p(x) = p_0(x) + p_1(x)$ is given by [318]

$$p(x) = \frac{x^{k_+/\gamma - 1}(1-x)^{k_-/\gamma - 1}}{B(k_+/\gamma, k_-/\gamma)}. \tag{6.3.6}$$

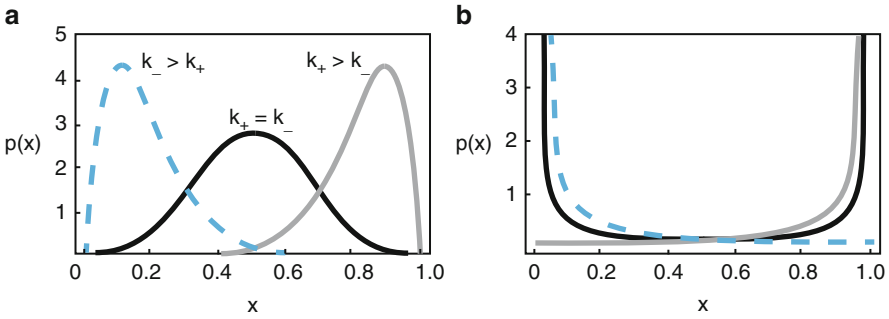


Fig. 6.4: Steady-state protein density $p(x)$ for a simple regulated network in which the promoter transitions between an active and inactive state at rates $k_\pm$. (**a**) Case $k_\pm/k > 1$: there is a graded density that is biased towards $x = 0, 1$ depending on the ratio k_+/k_-. (**b**) Case $k_\pm/k < 1$: there is a binary density that is concentrated around $x = 0, 1$ depending on the ratio k_+/k_-

In Fig. 6.4, we plot $p(x)$, $0 < x < 1$ for various values of $K_\pm = k_\pm/\gamma$. It can be seen that when the rates $k_\pm$ of switching between the active and inactive gene states are faster than the rate of degradation k, then the steady-state density is unimodal (graded), whereas if the rate of degradation is faster, then the density tends to be concentrated around $x = 0$ or $x = 1$, consistent with a binary process. In other words, if switching between promoter states is much slower than other processes, then one can have transcriptional contribution to protein bursting [318]. This scenario tends to occur in eukaryotic gene expression, for which the presence of nucleosomes and the packing of DNA–nucleosome complexes into chromatin generally make promoters inaccessible to the transcriptional machinery. Hence, transitions between open and closed chromatin structures, corresponding to active and repressed promoter states, can be quite slow.

Finally, note that a model identical in form to the above has also been applied to gene expression dynamics in a randomly varying environment [599]. In the latter case, x represents the concentration of mRNA and γ is the rate of degradation. The rate k of mRNA production takes on two values, depending on a binary-valued environmental input $n(t)$, with $k = k_0$ if $n(t) = 0$ and $k = k_1$ if $n(t) = 1$. The environment randomly switches between its two states at the rates $k_\pm$. Equations (6.3.2) thus become

$$\frac{\partial p_0}{\partial t} = -\frac{\partial}{\partial x}([k_0 - \gamma x]p_0(x,t)) + k_- p_1(x,t) - k_+ p_0(x,t) \tag{6.3.7a}$$

$$\frac{\partial p_1}{\partial t} = -\frac{\partial}{\partial x}([k_1 - \gamma x]p_1(x,t)) + k_+ p_0(x,t) - k_- p_1(x,t), \tag{6.3.7b}$$

where $p_j(x,t)$ is the probability density for mRNA concentration x given the environmental input is $n(t) = j, j = 0,1$. The analysis of the steady-state density proceeds as before and one finds [599]

$$p_0(x) = C(\gamma x - k_0)^{-1+k_+/\gamma}(k_1 - \gamma x)^{k_-/\gamma}, \quad p_1(x) = C(\gamma x - k_0)^{k_+/\gamma}(k_1 - \gamma x)^{-1+k_-/\gamma} \tag{6.3.8}$$

for some constant C. Imposing the normalization conditions, then determines C as

$$C = \gamma \left[(k_1 - k_0)^{(k_+ + k_-)/\gamma} B(k_+/\gamma, k_-/\gamma) \right]^{-1}.$$

It follows from the analog of Fig. 6.4 that if the mRNA degradation rate is faster than the rate of environmental fluctuations, then the steady-state density of mRNA tracks the state of the environment with $p(x)$ localized around $x = 0$ ($x = 1$) when $k_- > k_+$ ($k_+ > k_-$). Stochastic switching has been suggested as a survival strategy used by populations of yeast cells in fluctuating environments [1].

6.3.2 Protein Fluctuations and the Linear Noise Approximation

In the above analysis, we considered the distribution of proteins arising from a single gene, in which the only source of noise came from the random switching of the promoter. We now want to estimate the size of protein fluctuations in a population of $n_{\max}$ genes that takes into account intrinsic noise effects due to a finite number of proteins. Let n_1 denote the number of active genes and n_2 the number of proteins. Setting $x_j = \langle n_j \rangle / \Omega$, where Ω is the system size, the various reactions and the corresponding rate equations based on mass action (valid in the limit $\Omega \to \infty$) are as follows:

1. *Gene activation and inactivation*

$$n_1 \xrightarrow{k_+(n_{\max} - n_1)} n_1 + 1, \quad n_1 \xrightarrow{k_- n_1} n_1 - 1$$

with

$$\frac{dx_1}{dt} = k_+(x_{\max} - x_1) - k_- x_1.$$

2. *Protein production and degradation*

$$n_2 \xrightarrow{r n_1} n_2 + 1, \quad n_2 \xrightarrow{\gamma n_2} n_2 - 1$$

with

$$\frac{dx_2}{dt} = rx_1 - \gamma x_2.$$

In order to take into account the effects of intrinsic noise, it is necessary to turn to the associated master equation. Let $P = P(n_1, n_2, t)$ denote the probability that there are n_1 active genes and n_2 proteins at time t. Then

$$\begin{aligned}\frac{dP}{dt} &= k_+(n_{\max} - n_1 + 1)P(n_1 - 1, n_2, t) + k_-(n_1 + 1)P(n_1 + 1, n_2, t) \\ &\quad + rn_1 P(n_1, n_2 - 1, t) + \gamma(n_2 + 1)P(n_1, n_2 + 1, t) \\ &\quad - [k_+(n_{\max} - n_1) + k_- n_1 + rn_1 + \gamma n_2]\, P(n_1, n_2, t). \end{aligned} \tag{6.3.9}$$

Since the transition rates are linear in n_1 and n_2, one could determine the means and variances by taking moments. However, this method is not applicable to master equations with nonlinear transition rates. Therefore, we will follow the approximation method introduced in Sect. 3.2, whereby the master equation is reduced to a Fokker–Planck (FP) equation by carrying out a system-size expansion. The resulting FP equation can then be linearized about a stable fixed point of the deterministic rate equations, resulting in a multivariate OU process that can be used to calculate means and variances [162, 164, 625]. One of the useful features of the *linear-noise approximation* is that it can be applied systematically, once the mass-action kinetic equations are expressed in the general form (6.3.17) as described in Box 6A. For the given regulatory network, there are two chemical species ($N = 2$) and four single-step reactions ($R = 4$). For $a = 1, 2$ (gene activation and inactivation), we have $S_{i,1} = \delta_{i,1}, S_{i,2} = -\delta_{i,1}, f_1(\mathbf{x}) = k_+(x_{\max} - x_1)$, and $f_2(\mathbf{x}) = k_- x_1$. Expressing the master equation as (6.3.18) and carrying out a diffusion approximation then leads to the FP equation (6.3.20) with drift terms

$$V_1(\mathbf{x}) = k_+(x_{\max} - x_1) - k_- x_1, \quad V_2(\mathbf{x}) = rx_1 - \gamma x_2$$

and a diagonal diffusion matrix $\mathbf{D}$ with nonzero components

$$D_{11} = k_+(x_{\max} - x_1) + k_- x_1, \quad D_{22} = rx_1 + \gamma x_2.$$

In the deterministic limit, we recover the kinetic equations expressed as

$$\frac{dx_i}{dt} = V_i(\mathbf{x}).$$

It immediately follows that there is a unique fixed point given by

$$x_1^* = \frac{k_+}{k_+ + k_-} x_{\max}, \quad x_2^* = \frac{r}{\gamma} x_1^*.$$

Linearizing the corresponding Langevin equation about this fixed point by setting $X_i(t) = x_i^* + \Omega^{-1/2} Y_i(t)$ then yields the OU process (6.3.25) for Y_i, which takes the explicit form

$$dY_1 = -(k_+ + k_-)Y_1 dt + dW_1, \quad dY_2 = [rY_1 - \gamma Y_2]dt + dW_2, \tag{6.3.10}$$

with $W_1(t)$ and $W_2(t)$ independent Wiener processes satisfying

$$\langle dW_1(t) dW_1(t') \rangle = [k_+(x_{\max} - x_1^*) + k_- x_1^*]\delta(t-t')dtdt' = 2k_- x_1^* \delta(t-t')dtdt',$$
$$\langle dW_2(t) dW_2(t') \rangle = [rx_1^* + \gamma x_2^*]\delta(t-t')dtdt' = 2rx_1^* \delta(t-t')dtdt'.$$

Introducing the stationary covariance matrix

$$\Sigma_{ij} = \langle [Y_i(t) - \langle Y_i(t) \rangle][Y_j(t) - \langle Y_j(t) \rangle] \rangle$$

one sees that $Y_i(t)$ is a Gaussian process with zero mean and covariances determined from the matrix equation

$$\mathbf{A}\Sigma + \Sigma \mathbf{A}^T = -\mathbf{D}, \tag{6.3.11}$$

with

$$\mathbf{A} = \begin{pmatrix} -(k_+ + k_-) & 0 \\ r & -\gamma \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} 2k_- x_1^* & 0 \\ 0 & 2rx_1^* \end{pmatrix}.$$

Finally, solving the matrix equation (6.3.11) for the covariance gives the Fano factors (see Ex. 6.4):

$$\frac{\text{var}[n_1]}{\langle n_1 \rangle} = \frac{k_-}{k_+ + k_-} = 1 - \langle n_1 \rangle / n_{\max}, \tag{6.3.12a}$$

$$\frac{\text{var}[n_2]}{\langle n_2 \rangle} = 1 + \langle n_2 \rangle \frac{\gamma}{k_+ + k_- + \gamma} \frac{\text{var}[n_1]}{\langle n_1 \rangle^2}. \tag{6.3.12b}$$

Note that $\langle n_j \rangle = \Omega x_j^*$ and $\text{var}[n_j] = \Omega \Sigma_{jj}$. We immediately see that the presence of a transcription factor increases the Fano factor of the protein above one.

An alternative approach to analyzing the multivariate OU process derived from the linear noise approximation is to Fourier transform the corresponding multivariate Langevin equation (6.3.10) and calculate the spectrum of the protein concentration [328, 624]. First, since we are ultimately interested in protein number fluctuations, we rescale the Langevin equation by setting $\Delta n_j = \sqrt{\Omega} Y_j = n_j - \langle n_j \rangle$ and use the white noise formulation (see Sect. 2.2.5):

$$\frac{d\Delta n_1}{dt} = -(k_+ + k_-)\Delta n_1 + \eta_1, \quad \frac{d\Delta n_2}{dt} = r\Delta n_1 - \gamma \Delta n_2 + \eta_2, \tag{6.3.13}$$

with $\eta_1(t)$ and $\eta_2(t)$ independent Gaussian white noise processes satisfying

$$\langle \eta_1(t)\eta_1(t')\rangle = [k_+(n_{\max} - \langle n_1\rangle) + k_-\langle n_1\rangle]\delta(t-t') = 2k_-\langle n_1\rangle\delta(t-t'),$$
$$\langle \eta_2(t)\eta_2(t')\rangle = [r\langle n_1\rangle + \gamma\langle n_2\rangle]\delta(t-t') = 2r\langle n_1\rangle\delta(t-t').$$

Fourier transforming the linear equations (6.3.13) with

$$\Delta n_j(t) = \int_{-\infty}^{\infty} \widetilde{\Delta n}_j(\omega) \mathrm{e}^{-i\omega t}\frac{d\omega}{2\pi}$$

yields

$$-i\omega\widetilde{\Delta n}_1 = -(k_+ + k_-)\widetilde{\Delta n}_1 + \widetilde{\eta_1}, \quad -i\omega\widetilde{\Delta n}_2 = r\widetilde{\Delta n}_1 - \gamma\widetilde{\Delta n}_2 + \tilde{\eta}_2. \tag{6.3.14}$$

It follows that

$$\begin{aligned}\widetilde{\Delta n}_2 &= \frac{r\widetilde{\Delta n}_1}{\gamma - i\omega} + \frac{\tilde{\eta}_2}{\gamma - i\omega} \\ &= \frac{r\tilde{\eta}_1}{(k_+ + k_- - i\omega)(\gamma - i\omega)} + \frac{\tilde{\eta}_2}{\gamma - i\omega}.\end{aligned}$$

From the spectral analysis of Sect. 2.2.5, we have

$$\langle \tilde{\eta}_1(\omega)\tilde{\eta}_1(\omega')\rangle = 2k_-\langle n_1\rangle \cdot 2\pi\delta(\omega+\omega'), \quad \langle \tilde{\eta}_2(\omega)\tilde{\eta}_2(\omega')\rangle = 2r\langle n_1\rangle \cdot 2\pi\delta(\omega+\omega').$$

Hence, the spectrum of the protein fluctuations, defined by $\langle \widetilde{\Delta n}_2(\omega)\widetilde{\Delta n}_2(\omega')\rangle = S_2(\omega)\delta(\omega+\omega')$, is

$$S_2(\omega) = \frac{r^2(2k_-\langle n_1\rangle)}{(\omega^2 + (k_+ + k_-)^2)(\omega^2 + \gamma^2)} + \frac{2r\langle n_1\rangle}{\omega^2 + \gamma^2}. \tag{6.3.15}$$

It follows that

$$\begin{aligned}\mathrm{var}[n_2] = \langle (\Delta n_2)^2\rangle &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \langle \widetilde{\Delta n}_2(\omega)\widetilde{\Delta n}_2(\omega')\rangle \mathrm{e}^{-i\omega t}\mathrm{e}^{-i\omega' t}\frac{d\omega}{2\pi}\frac{d\omega'}{2\pi} \\ &= \int_{-\infty}^{\infty} S_2(\omega)\frac{d\omega}{2\pi}.\end{aligned}$$

The integral can be evaluated using partial fractions and the identity

$$\int_{-\infty}^{\infty} \frac{d\omega}{\omega^2 + a^2} = \frac{\pi}{a},$$

which gives

$$\begin{aligned}\text{var}[n_2] &= \frac{r\langle n_1\rangle}{\gamma} + \frac{r^2 k_- \langle n_1\rangle}{(k_+ + k_-)^2 - \gamma^2}\left(\frac{1}{\gamma} - \frac{1}{k_+ + k_-}\right) \\ &= \frac{r\langle n_1\rangle}{\gamma} + \frac{r^2\langle n_1\rangle/\gamma}{k_+ + k_- + \gamma}\frac{k_-}{(k_+ + k_-)} \\ &= \langle n_2\rangle + (r/\gamma)^2 \frac{\gamma}{k_+ + k_- + \gamma}\langle n_1\rangle(1 - \langle n_1\rangle/n_{\max}). \end{aligned} \tag{6.3.16}$$

This agrees with Eq. (6.3.12b). Finally, note that there are two contributions to the size of protein fluctuations. First, there is the output noise $\langle n_2\rangle$ arising from the production of a finite number of proteins in which the variance equals the mean, reflecting a pure Poisson process. The second contribution arises from the random switching of the promoter and is proportional to the binomial variance $p_1(1-p_1)$ where $p_1 = \langle n_1\rangle/n_{\max}$ is the mean fraction of active genes. For further applications of frequency domain analysis to feedforward gene networks see Exs. 6.5 and 6.6.

Box 6A. Linear noise approximation.

Suppose that the mass-action kinetics of a general biochemical or gene network is written in the form

$$\frac{dx_i}{dt} = \sum_{a=1}^{R} S_{ia} f_a(\mathbf{x}), \quad i = 1,\ldots,N \tag{6.3.17}$$

where a labels a single-step reaction and $\mathbf{S}$ is the so-called $N \times R$ stochiometric matrix for N molecular species and R reactions. Thus S_{ia} specifies the change in the number of molecules of species i in a given reaction a. The functions f_a are known transition intensities or *propensities*. Given this notation, the corresponding master equation is

$$\frac{dP(\mathbf{n},t)}{dt} = \Omega \sum_{a=1}^{R}\left(\prod_{i=1}^{N}\mathbb{E}^{-S_{ia}} - 1\right) f_a(\mathbf{n}/\Omega)P(\mathbf{n},t), \tag{6.3.18}$$

where Ω represents the system size. Typically, Ω is the volume of the well-mixed compartment where reactions occur or the total number of molecules in cases where there is number conservation. Here $\mathbb{E}^{-S_{ia}}$ is a step or ladder operator such that for any function $g(\mathbf{n})$,

$$\mathbb{E}^{-S_{ia}} g(n_1,\ldots,n_i,\ldots,n_N) = g(n_1,\ldots,n_i - S_{ia},\ldots,n_N). \tag{6.3.19}$$

A diffusion approximation of the master equation can now be obtained along identical lines to Sect. 3.2 (see also [162]). That is, set $f_a(\mathbf{n}/\Omega)P(\mathbf{n},t) \to f_a(\mathbf{x})\mathbf{p}(\mathbf{x},t)$ and use the fact that

$$\prod_{i=1}^{N} \mathbb{E}^{-S_{ia}} h(\mathbf{x}) = h(\mathbf{x} - \mathbf{S}_a/\Omega)$$
$$= h(\mathbf{x}) - \Omega^{-1} \sum_{i=1}^{N} S_{ia} \frac{\partial h}{\partial x_i} + \frac{1}{\Omega^2} \sum_{i,j=1}^{N} S_{ia} S_{ja} \frac{\partial^2 h(\mathbf{x})}{\partial x_i \partial x_j} + O(\Omega^{-3}).$$

Carrying out a Taylor expansion of the master equation to second order thus yields the multivariate FP equation

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{N} \frac{\partial V_i(\mathbf{x}) p(\mathbf{x},t)}{\partial x_i} + \frac{1}{2\Omega} \sum_{i,j=1}^{N} \frac{\partial^2 D_{ij}(\mathbf{x}) p(\mathbf{x},t)}{\partial x_i \partial x_j}, \tag{6.3.20}$$

where

$$V_i(\mathbf{x}) = \sum_{a=1}^{R} S_{ia} f_a(\mathbf{x}), \quad D_{ij}(\mathbf{x}) = \sum_{a=1}^{R} S_{ia} S_{ja} f_a(\mathbf{x}). \tag{6.3.21}$$

The FP equation (6.3.20) corresponds to the multivariate Langevin equation

$$dX_i = V_i(\mathbf{X})dt + \frac{1}{\sqrt{\Omega}} \sum_{a=1}^{R} B_{ia}(\mathbf{X}) dW_a(t), \tag{6.3.22}$$

where $W_a(t)$ are independent Wiener processes and $\mathbf{D} = \mathbf{B}\mathbf{B}^T$, that is,

$$B_{ia} = S_{ia} \sqrt{f_a(\mathbf{x})}. \tag{6.3.23}$$

Now suppose that the deterministic system, written as

$$\frac{dx_i}{dt} = V_i(\mathbf{x}),$$

has a unique stable fixed point $\mathbf{x}^*$ for which $V_i(\mathbf{x}^*) = 0$ and introduce the Jacobian matrix $\mathbf{A}$ with

$$A_{ij} = \left. \frac{\partial V_i}{\partial x_j} \right|_{\mathbf{x}=\mathbf{x}^*}. \tag{6.3.24}$$

The Langevin equation suggests that, after a transient phase, the stochastic dynamics is characterized by Gaussian fluctuations about the fixed point. Substituting $X_i(t) = x_i^* + Y_i(t)/\sqrt{\Omega}$ into the Langevin equation (6.3.20) and keeping only lowest order terms in $\Omega^{-1/2}$ yields the Ornstein–Uhlenbeck (OU) process

$$dY_i = \sum_{j=1}^{N} A_{ij} Y_j dt + \sum_{a=1}^{R} B_{ia}(\mathbf{x}^*) dW_a(t). \tag{6.3.25}$$

Introducing the stationary covariance matrix

$$\Sigma_{ij} = \langle [Y_i(t) - \langle Y_i(t) \rangle][Y_j(t) - \langle Y_j(t) \rangle] \rangle$$

it immediately follows from the analysis of the multivariate OU process (see Ex. 2.7), that

$$\mathbf{A}\Sigma + \Sigma \mathbf{A}^T = -\mathbf{B}\mathbf{B}^T. \tag{6.3.26}$$

6.3.3 Autoregulatory Network

So far we have considered a simple feedforward regulatory network. However, much of the complexity in gene networks arises from feedback, in which proteins influence their own synthesis directly or indirectly by acting as transcription factors within a regulatory network. A common example is autoregulation, in which a gene is directly regulated by its own gene product [625] (see Fig. 6.5a). A simple kinetic model of negative autoregulatory feedback is

$$\frac{dx_1}{dt} = -\gamma x_1 + F(x_2), \quad \frac{dx_2}{dt} = r x_1 - \gamma_p x_2, \tag{6.3.27}$$

where $x_1(t)$ and $x_2(t)$ denote the concentrations (or number) of mRNA and protein molecules at time t. The parameters γ, γ_p represent the degradation rates, r represents the translation rate of proteins, and $F(y)$ represents the nonlinear feedback effect of the protein on the transcription of mRNA. A typical choice for F in the case of a repressor is the Hill function

$$F(y) = \frac{k}{1 + (y/K)^n}. \tag{6.3.28}$$

We will assume that the network acts in a regime where the Hill function is approximately linear with $F(y) = k_0 - ky$. The analysis of intrinsic noise proceeds along similar lines to regulated gene transcription.

Let $P = P(m,n,t)$ denote the probability that there are m mRNA and n proteins at time t. Then

$$\begin{aligned} \frac{dP}{dt} &= \Omega k_0 P(m-1,n,t) + [kn + \gamma(m+1)]P(m+1,n,t) \\ &\quad + rmP(m,n-1,t) + \gamma_p(n+1)P(m,n+1,t) \\ &\quad - [\Omega k_0 + (kn + \gamma m) + rm + \gamma_p n] P(m,n,t). \end{aligned} \tag{6.3.29}$$

In order to carry out a linear noise approximation, we first rewrite the kinetic equations in the general form (6.3.17) with two chemical species ($N=2$) and four single-step reactions ($R=4$). For $a=1,2$ (mRNA production and degradation/repression), we have $S_{i,1}=\delta_{i,1}, S_{i,2}=-\delta_{i,1}, f_1(\mathbf{x})=k_0$, and $f_2(\mathbf{x})=kx_2+\gamma x_1$. Expressing the master equation as (6.3.18) and carrying out a diffusion approximation then leads to the FP equation (6.3.20) with drift terms

$$V_1(\mathbf{x})=k_0-kx_2-\gamma x_1, \quad V_2(\mathbf{x})=rx_1-\gamma_p x_2$$

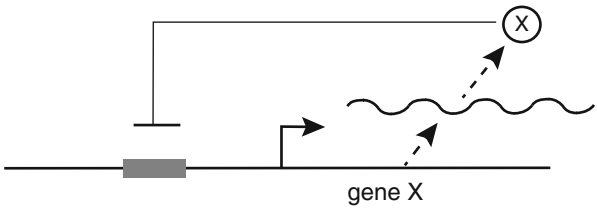


Fig. 6.5: Negative autoregulatory network. A gene X is repressed by its own protein product

and a diagonal diffusion matrix $\mathbf{D}$ with nonzero components

$$D_{11}=k_0+kx_2+\gamma x_1, \quad D_{22}=rx_1+\gamma_p x_2.$$

There is a unique fixed point of the deterministic dynamics (in the linear regime)

$$x_1^*=\frac{k_0\gamma_p}{\gamma\gamma_p+kr}, \quad x_2^*=\frac{r}{\gamma_p}x_1^*.$$

Linearizing the corresponding Langevin equation about this fixed point by setting $X_i(t)=x_i^*+\Omega^{-1/2}Y_i(t)$ then yields the OU process (6.3.25) for Y_i. Introducing the stationary covariance matrix

$$\Sigma_{ij}=\langle[Y_i(t)-\langle Y_i(t)\rangle][Y_j(t)-\langle Y_j(t)\rangle]\rangle$$

one sees that $Y_i(t)$ is a Gaussian process with zero mean and covariances determined from the matrix equation

$$\mathbf{A}\Sigma+\Sigma\mathbf{A}^T=-\mathbf{D} \tag{6.3.30}$$

with

$$\mathbf{A}=\begin{pmatrix}-\gamma & -k\\ r & -\gamma_p\end{pmatrix}, \quad \mathbf{D}=\begin{pmatrix}kx_2^*+\gamma x_1^* & 0\\ 0 & rx_1^*+\gamma_p x_2^*\end{pmatrix}.$$

Solving the matrix equation (6.3.30) yields (see Ex. 6.7)

$$\Sigma_{12} = \Sigma_{21} = \frac{\eta}{1+\eta}\left(1 - \frac{\phi}{1+b\phi}\right)x_2^*, \quad \Sigma_{22} = x_2^* + \frac{r}{\gamma_p}\Sigma_{12},$$

where

$$b = \frac{r}{\gamma}, \quad \eta = \frac{\gamma_p}{\gamma}, \quad \phi = \frac{k}{\gamma_p}.$$

Here b is the burst size, η is the ratio of degradation rates, and ϕ describes the strength of the negative feedback. It follows that the Fano factor for proteins is

$$\frac{\text{var}[n]}{\langle n \rangle} = 1 + \frac{b}{1+\eta}\left(1 - \frac{\phi}{1+b\phi}\right). \tag{6.3.31}$$

The above analysis establishes the negative feedback can reduce fluctuations in protein number (see Fig. 6.2b). That is, in the absence of feedback ($\phi = 0$), the Fano factor is $1 + b/(1+\eta)$, which is clearly larger than the case $\phi > 0$. Also note that when $\eta \ll 1$ and $b \gg 1$, we recover the result obtained from the protein translation model of Sect. 6.2.

6.4 Genetic Switches and Oscillators

Once feedback and nonlinearities are included in gene networks, a rich repertoire of dynamics can occur. Here we briefly consider two important classes of dynamical gene networks, namely, switches and oscillators.

6.4.1 Mutual Repressor Model of a Genetic Switch

Considerable insight into genetic switches has been obtained by constructing a synthetic version of a switch in *E. coli*, in which the gene product of the switch is a fluorescent reporter protein [206]. This allows the flipping of the switch to be observed by measuring the fluorescent level of the cells. The underlying gene circuit is based on a mutual repressor model (see Fig. 6.6). It consists of two repressor proteins whose transcription is mutually regulated. That is, the protein product of one gene binds to the promoter of the other gene and represses its output. For simplicity, the explicit dynamics of transcription and translation are ignored so that we only model the mutual effects of the proteins on protein production. Denoting the concentrations of the proteins by $x(t), y(t)$, the resulting kinetic equations are

$$\frac{dx}{dt} = -\gamma x + \frac{r}{1+Ky^n}, \quad \frac{dy}{dt} = -\gamma y + \frac{r}{1+Kx^n}. \tag{6.4.1}$$

Here γ is the rate of protein degradation, r is the rate of protein production in the absence of repression, and K is a binding constant for the repressors. As in the model of autoregulation, negative feedback is modeled in terms of a Hill function with Hill coefficient n. It is convenient to rewrite the equations in nondimensional form by measuring x and y in units of $K^{-1/n}$ and time in units of γ^{-1}:

$$\frac{du}{dt} = -u + \frac{\alpha}{1+v^n}, \quad \frac{dv}{dt} = -v + \frac{\alpha}{1+u^n}, \tag{6.4.2}$$

with $\alpha = rK^{1/n}/\gamma$. Analysis of the fixed point solutions of this pair of equations establishes that the mutual repressor model acts as a bistable switch. For simplicity, consider the case $n = 2$ (protein dimerization). The fixed point equation for u is

$$u = \alpha \left[1 + \left(\frac{\alpha}{1+u^2} \right)^2 \right]^{-1},$$

which can be rearranged to yield a product of two polynomials:

$$(u^2 - \alpha u + 1)(u^3 + u - \alpha) = 0.$$

The cubic is a monotonically increasing function of u and thus has a single root given implicitly by

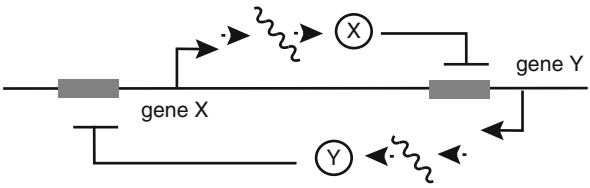


Fig. 6.6: Mutual repressor model of a genetic switch. A gene X expresses a protein X that represses the transcription of gene Y and the protein Y represses the transcription of gene X

$$u = \frac{\alpha}{1+u^2} = v.$$

This solution is guaranteed by the exchange symmetry of the underlying equations. The roots of the quadratic are given by

$$u = U_{\pm} \equiv \frac{1}{2}\left[\alpha \pm \sqrt{\alpha^2 - 4} \right],$$

with $v = U_{\mp}$. It immediately follows that there is a single fixed point when $\alpha < 2$ and three fixed points when $\alpha > 2$. Moreover, linear stability analysis establishes that the symmetric solution is stable when $\alpha < 2$ and undergoes a pitchfork bifurcation at the critical value $\alpha_c = 2$ where it becomes unstable and a pair of stable fixed points emerge.

Given that the deterministic system is bistable, one can now investigate the effects of intrinsic noise by constructing a master equation along the lines of Sect. 6.3. We will construct the master equation for a slightly simplified mutual repressor model consisting of a single promoter site; if a dimer of one protein is bound to the site then this represses the expression of the other [328, 470]. Thus the promoter can be in three states O_j, $j = 0,1,2$: no dimer is bound to the promoter (O_0); a dimer of protein X is bound to the promoter (O_1); a dimer of protein Y is bound to the promoter (O_2). Suppose that the number of proteins X and Y are n and m, respectively. The state transition diagram for the three promoter states is then

$$O_1 \underset{n(n-1)\kappa}{\overset{\beta\kappa}{\rightleftharpoons}} O_0 \underset{\beta\kappa}{\overset{m(m-1)\kappa}{\rightleftharpoons}} O_2,$$

where κ is a rate and β is a nondimensional dissociation constant. Protein X (Y) is produced at a rate α when the promoter is in the states $O_{0,1}$ ($O_{0,2}$), and both proteins are degraded at a rate γ in all three states. Let $p_j(n,m,t)$, $j = 0,1,2$, be the probability that there are n (m) proteins X (Y) and the promoter is in state j at time t. The master equation for $\mathbf{p} = (p_0, p_1, p_2)^T$ is given by

$$\frac{d}{dt}p_j(n,m,t) = \sum_{j=0,1,2}\sum_{n',m'}\left[\delta_{n,n'}\delta_{m,m'}A_{jk} + \delta_{j,k}W^j_{nm,n'm'}\right]p_k(n',m',t), \tag{6.4.3}$$

where

$$\mathbf{A} = \kappa\begin{pmatrix} -n(n-1)-m(m-1) & \beta & \beta \\ n(n-1) & -\beta & 0 \\ m(m-1) & 0 & -\beta \end{pmatrix}, \tag{6.4.4}$$

and

$$\begin{aligned}&\sum_{n',m'} W^0_{nm,n'm'}p_0(n',m',t)\\ &= \gamma[(n+1)p_0(n+1,m,t) + (m+1)p_0(n,m+1,t) - (n+m)p_0(n,m,t)]\\ &\quad + \alpha(p_0(n-1,m,t) + p_0(n,m-1,t) - 2p_0(n,m,t))\end{aligned} \tag{6.4.5a}$$

$$\begin{aligned}&\sum_{n',m'} W^1_{nm,n'm'}p_1(n',m',t)\\ &= \gamma[(n+1)p_1(n+1,m,t) + (m+1)p_1(n,m+1,t) - (n+m)p_1(n,m,t)]\\ &\quad + \alpha(p_1(n-1,m,t) - p_1(n,m,t))\end{aligned} \tag{6.4.5b}$$

$$\begin{aligned}&\sum_{n',m'} W^2_{nm,n'm'}p_1(n',m',t)\\ &= \gamma[(n+1)p_2(n+1,m,t) + (m+1)p_2(n,m+1,t) - (n+m)p_2(n,m,t)]\\ &\quad + \alpha(p_2(n,m-1,t) - p_2(n,m,t)).\end{aligned} \tag{6.4.5c}$$

Kepler and Elston [328] consider two approximations of the master equation, one based on a system-size expansion of the W^j terms with respect to the mean number $N = \alpha/\gamma$ of proteins when the promoter is in state O_0 and the other based on a QSS approximation. The latter assumes that the rates of protein production and degradation are much slower than the rates of switching between promoter states. First, introduce the rescaling $t \to t\gamma$ and set $x = n/N$, $y = m/N$. The master equation for the resulting probability densities $p_j(x,y,t)$ takes the form

$$\frac{\partial}{\partial t} p_j(x,y,t) = \sum_{j=0,1,2} \left[\frac{1}{\varepsilon} A_{jk} + N\delta_{j,k} \mathbb{W}^j \right] p_k(x,y,t), \tag{6.4.6}$$

where $\varepsilon = \gamma^3/\kappa\alpha^2$ and $b = \beta\gamma^2/\alpha^2$ are dimensionless parameters,

$$\mathbf{A} = \begin{pmatrix} -x(x-1/N) - y(y-1/N) & b & b \\ x(x-1/N) & -b & 0 \\ y(y-1/N) & 0 & -b \end{pmatrix}, \tag{6.4.7}$$

and $\mathbb{W}^j$ are differential shift operators

$$\mathbb{W}^0 = \left(\mathrm{e}^{\partial_x/N} - 1 \right) x + \left(\mathrm{e}^{\partial_y/N} - 1 \right) y + \left(\mathrm{e}^{-\partial_x/N} + \mathrm{e}^{-\partial_y/N} - 2 \right) \tag{6.4.8a}$$

$$\mathbb{W}^1 = \left(\mathrm{e}^{\partial_x/N} - 1 \right) x + \left(\mathrm{e}^{\partial_y/N} - 1 \right) y + \left(\mathrm{e}^{-\partial_x/N} - 1 \right) \tag{6.4.8b}$$

$$\mathbb{W}^2 = \left(\mathrm{e}^{\partial_x/N} - 1 \right) x + \left(\mathrm{e}^{\partial_y/N} - 1 \right) y + \left(\mathrm{e}^{-\partial_y/N} - 1 \right). \tag{6.4.8c}$$

The latter are a way of representing a Taylor expansion. That is, for any smooth function $f(x)$,

$$\begin{aligned} f(x \pm \Delta x) &= f(x) \pm f'(x)\Delta x + f''(x)\Delta x^2/2! \pm \dots \\ &= \left(1 \pm \Delta x \partial_x + \frac{\Delta x^2}{2!} \partial_x^2 \pm \dots \right) f(x) = \mathrm{e}^{\pm \Delta x \partial_x} f(x). \end{aligned}$$

If the promoter transitions are fast and the expected number of protein molecules is large, then there are two small parameters in the model, ε and $1/N$. Taking the limits $\varepsilon \to 0$ and $N \to \infty$ in either order, one obtains the kinetic equations (see also Ex. 6.8)

$$\frac{dx}{dt} = f(x,y), \quad \frac{dy}{dt} = f(y,x), \quad \text{with } f(x,y) = \frac{1}{1 + \frac{y^2}{b+x^2}} - x. \tag{6.4.9}$$

One finds that the deterministic system is bistable for $0 < b < b_c = 4/9$ (see Fig. 6.7). At the critical point $b = b_c$ there is a saddle-node bifurcation in which a stable/unstable pair annihilate so that for $b > b_c$ there is a single stable fixed point. There are then two approximations of the full master equation that can be used to explore the effects of noise-induced transitions in the bistable regime, depending on

whether one considers the system-size expansion in $1/N$ for fixed ε or the QSS expansion in ε for fixed N. For the sake of illustration, we focus on the former. Taylor expanding the differential operators $\mathbb{W}^j$ and keeping only the leading order terms yields the multivariate differential Chapman–Kolmogorov (CK) equation [328, 470]

$$\frac{\partial p_j}{\partial t} = -\frac{\partial F_j(x)p_j}{\partial x} - \frac{\partial G_j(y)p_j}{\partial y} + \frac{1}{\varepsilon}\sum_{k=0,1,2} A_{jk}(x,y)p_k \tag{6.4.10}$$

with

$$\begin{aligned} F_0(x) &= 1-x, \quad F_1(x) = 1-x, \quad F_2(x) = -x \\ G_0(y) &= 1-y, \quad G_1(y) = -y, \quad G_2(y) = 1-y, \end{aligned} \tag{6.4.11}$$

and

$$\mathbf{A} = \begin{pmatrix} -x^2-y^2 & b & b \\ x^2 & -b & 0 \\ y^2 & 0 & -b \end{pmatrix}. \tag{6.4.12}$$

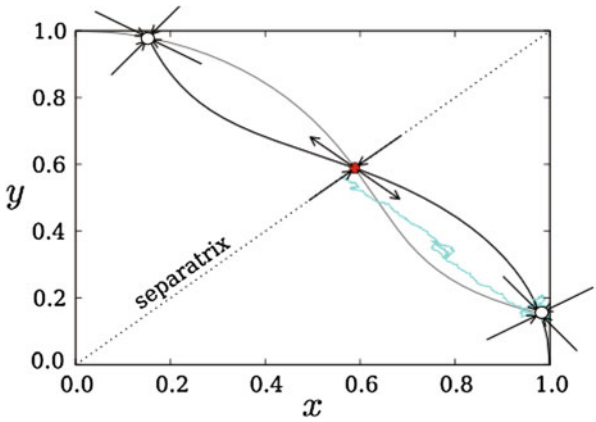


Fig. 6.7: Phase-plane dynamics of mutual repressor model analyzed by Kepler and Elston [328] and Newby [470] with $b = 0.15$. The *black curve* shows the y-nullcline and the *gray curve* shows the x-nullcline. The *open circles* show the stable fixed points; the *filled circle* shows the unstable saddle. The *irregular curve* shows a stochastic trajectory leaving the lower basin of attraction to reach the separatrix

The CK equation (6.4.10) describes an effective stochastic hybrid system in which the concentration of proteins X and Y play the role of the piecewise deterministic continuous variables, and the state of the promoter is the discrete variable that evolves according to a continuous-time Markov process. We have previously encountered stochastic hybrid systems in our analysis of voltage-gated ion channels (Sect. 3.5). One could now use a QSS approximation to obtain a Fokker–Planck

(FP) equation for the total probability density $p(x,y,t) = \sum_{j=0,1,2} p_j(x,y,t)$ along the lines outlined in Sect. 7.4 (see also Kepler and Elston [328]). However, a diffusion approximation of the full master equation based on an FP representation can generate exponential errors in the mean time of noise-induced escape from the basin of attraction of one of the metastable fixed points (see also Sects. 3.4 and 3.5). A more accurate estimate can be obtained using large deviation theory and the WKB methods outlined in Chap. 10, as has been shown for the mutual repressor model by Newby [470].

6.4.2 *The* lac *Operon*

The idea of a genetic switch was first proposed over 40 years ago by Jacob and Monod [296], in their study of the *lac* operon. When there is an abundance of glucose, *E. coli* uses glucose exclusively as a food source irrespective of whether or not other sugars are present. However, when glucose is unavailable, *E. coli* can feed on other sugars such as lactose, and this occurs via the *lac* operon switch that induces the expression of various genes. A variety of mathematical models of the *lac* operon

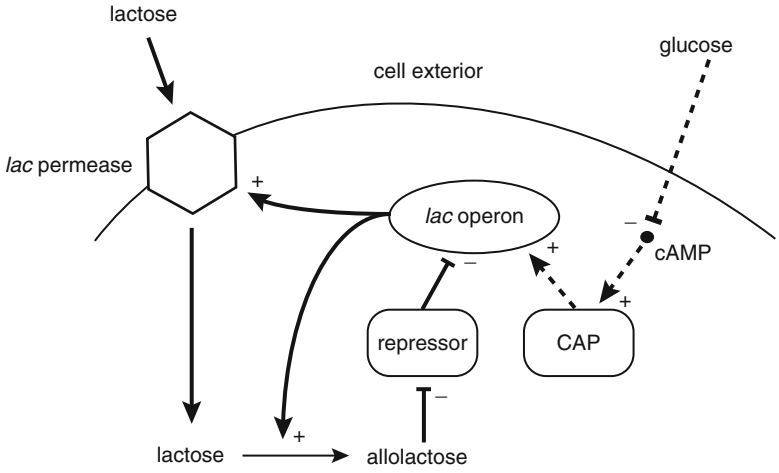


Fig. 6.8: Feedback control circuit of the *lac* operon. See text for details

have been developed over the years [239, 240, 557, 688, 693, 694]. Here we briefly describe a simplified model presented in Chap. 10 of Keener and Sneyd [322]. The basic feedback control mechanism is illustrated in Fig. 6.8. There are two control sites on the *lac* operon: (see Fig. 6.9), a repressor site that blocks RNAP from binding to the promoter site and a preceding control site to which a dimeric catabolic activator protein (CAP) molecule can bind provided it forms a complex with cyclic AMP (cAMP). Bound CAP promotes the binding of RNAP to the promoter region.

When there is sufficient glucose in the cell exterior, the action of cAMP is inhibited so that CAP cannot bind and the *lac* operon is repressed. On the other hand, when glucose is removed, the CAP–cAMP complex can bind to the activator site and activate the *lac* operon. The latter consists of several genes that code for the proteins responsible for lactose metabolism. One of these proteins is *lac* permease, which allows the entry of lactose into the cell that is enhanced by a positive feedback loop. The feedback mechanism involves another protein, β-galactosidase, which converts lactose into allolactose. Allolactose can bind to the repressor protein and prevent its binding to the repressor binding site. This further activates the *lac* operon, resulting in the further production of allolactose and increased entry of lactose via the *lac* permease.

Suppose that the CAP dynamics is ignored, so that we can focus on the positive feedback loop indicated in Fig. 6.8 by solid arrows. Let A denote the concentration of allolactose and similarly for lactose (L), the permease (P), the protein product β-galactosidase (B), mRNA (M), and the repressor (R). Let p_{on} and p_{off} denote the probabilities that the operon is on and off, respectively, with $p_{\text{on}} + p_{\text{off}} = 1$. Ignoring the effect of the CAP site, we have the simple kinetic scheme

$$\frac{dp_{\text{on}}}{dt} = k_{-r}(1 - p_{\text{on}}) - k_r R^* p_{\text{on}},$$

where R^* is the concentration of repressor in the activated state. Each activated repressor protein interacts with two molecules of allolactose to become inactivated, so from mass-action kinetics,

$$\frac{dR^*}{dt} = k_{-a}R - k_a A^2 R^*,$$

where the binding/unbinding of a single repressor molecule to the operon has a negligible effect on the total concentration $R_T = R + R^*$. The next simplification is to take these reactions to be much faster than those associated with gene expression so that p_{on} and R^* take the steady-state values

$$R^* = \frac{R_T}{1 + K_a A^2}, \quad p_{\text{on}} = \frac{1}{1 + K_r R^*},$$

with $K_a = k_a/k_{-a}$ and $K_r = k_r/k_{-r}$. Combining these two results gives the steady-state probability

$$p_{\text{on}} = \frac{1 + K_a A^2}{1 + K_r R_T + K_a A^2} \equiv \Gamma(A).$$

It follows that the concentration of mRNA is determined by the equation

$$\frac{dM}{dt} = \alpha_M \Gamma(A) - \gamma_M M, \tag{6.4.13a}$$

where α_M and γ_M are the rates of mRNA production and degradation. This is the first of the model equations. The next two equations represent the dynamics of the enzymes directly produced by the on-state of the operon, namely, permease and β-galactosidase:

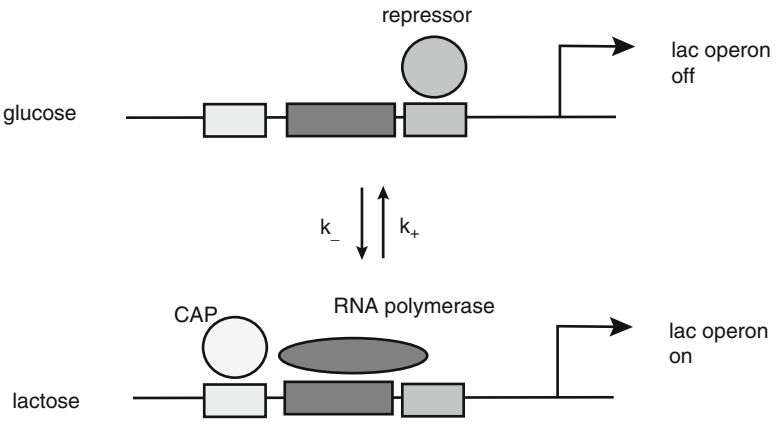


Fig. 6.9: Repressor and CAP sites for the *lac* operon

$$\frac{dP}{dt} = \alpha_P M - \gamma_P P, \tag{6.4.13b}$$

$$\frac{dB}{dt} = \alpha_B M - \gamma_B B. \tag{6.4.13c}$$

Note that although both enzymes are produced by different parts of the same mRNA, the effective production rates differ due to different times of production (permease is produced after β-galactosidase) and the time delay associated with permease migrating to the cell membrane. The final two equations specify the dynamics of lactose and allolactose based on Michaelis–Menten kinetics (see Box 6B). Let L_e be a fixed concentration of lactose exterior to the cell. Lactose enters the cell at a Michaelis–Menten rate proportional to the permease concentration P, where it is converted to allolactose via the enzymatic action of β-galactosidase; the latter also breaks down allolactose into glucose and galactose. Thus

$$\frac{dL}{dt} = \alpha_L P \frac{L_e}{K_{Le} + L_e} - \alpha_A B \frac{L}{K_L + L} - \gamma_L L \tag{6.4.13d}$$

$$\frac{dA}{dt} = \alpha_A B \frac{L}{K_L + L} - \beta_A B \frac{A}{K_A + A} - \gamma_A A. \tag{6.4.13e}$$

Keener and Sneyd [322] show that for physiologically based parameter values, the system of Eq. (6.4.13) exhibits bistability in the interior lactose concentration as a function of the exterior lactose concentration L_e. Note that the stochastic analysis

outlined in Sect. 6.4.1 for the mutual repressor model could be extended to the more complicated model of the *lac* operon in order to investigate the effects of intrinsic noise on the bistable switch.

6.4.3 Genetic Oscillator Network

There are numerous examples of gene circuits that support oscillations. Here we consider a relaxation oscillator consisting of an activator that increases its own production and that of a repressor, which in turn represses the production of the activator (see Fig. 6.10). Let x denote the concentration of the activator and y denote the concentration of the repressor. The resulting kinetic equations take the form

$$\frac{dx}{dt} = -\gamma_x x + r_{0x}\frac{1}{1+(x/K_d)^2+(y/K_d)^2} + r_x\frac{(x/K_d)^2}{1+(x/K_d)^2+(y/K_d)^2} \tag{6.4.14a}$$

$$\frac{dy}{dt} = -\gamma_y y + r_{0y}\frac{1}{1+(x/K_d)^2} + r_y\frac{(x/K_d)^2}{1+(x/K_d)^2}, \tag{6.4.14b}$$

where γ_x, γ_y are the degradation rates of the two proteins, r_{0x}, r_{0y} are protein production rates when respective promoters are not bound by transcription factors, and r_x, r_y are the enhanced production rates when the promoter sites are activated. (It is assumed that when the promoter of gene X is repressed, production of protein X is blocked.) The production terms are based on the equilibrium binding probabilities of the X and Y promoter domains. The Hill coefficient $n = 2$ arises because the transcription factors bind as dimers. In the case of the unbound promoter A of activator gene X, the binding reactions are

$$2X + A \rightleftharpoons A^+, \quad 2Y + A \rightleftharpoons A^-,$$

where $A^\pm$ denote the activated and repressed promoter states. In terms of the equilibrium law of mass action (see Sects. 1.4 and 4.1), the concentrations of the various reactants and products satisfy

$$\frac{[A^+]}{[X^2][A]} = \frac{1}{K_d}, \quad \frac{[A^-]}{[Y^2][A]} = \frac{1}{K_d},$$

with the dissociation constant K_d taken to be the same for both binding reactions. Denoting the total concentration of promoter domains of gene X by $T_A = [A] + [A^+] + [A^-]$, we have

$$\frac{[A^+]}{T_A - [A^+] - [A^-]} = \frac{[X^2]}{K_d}, \quad \frac{[A^-]}{T_A - [A^+] - [A^-]} = \frac{[Y^2]}{K_d}.$$

These can be solved to give the equilibrium probabilities that the X promoter domain is activated or repressed:

$$\frac{[A^+]}{T_A} = \frac{([X]/K_d)^2}{1+([X]/K_d)^2+([Y]/K_d)^2}, \quad \frac{[A^-]}{T_A} = \frac{([Y]/K_d)^2}{1+([X]/K_d)^2+([Y]/K_d)^2}.$$

A similar analysis of the single binding reaction

$$2X + R \rightleftharpoons R^+,$$

where R, R^+ are the unbound and activated states of the Y gene promoter, yields

$$\frac{[R^+]}{T_R} = \frac{([X]/K_d)^2}{1+([X]/K_d)^2},$$

where T_R is the total concentration of the Y gene promoter.

As in the case of the genetic switch, it is useful to nondimensionalize the equations by taking time to be in units of γ_y^{-1} and concentrations in units of K_d:

$$\frac{dx}{dt} = -\gamma x + \frac{R_{0x} + R_x x^2}{1+x^2+y^2} \tag{6.4.15a}$$

$$\frac{dy}{dt} = -y + \frac{R_{0y} + R_y x^2}{1+x^2}, \tag{6.4.15b}$$

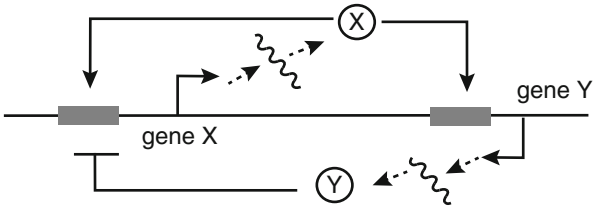


Fig. 6.10: Activator–repressor model of a genetic oscillator. A gene X expresses a protein X that activates the transcription of genes X and Y and protein Y represses the transcription of gene X

where $\gamma = \gamma_x/\gamma_y$ and $R_{0x} = r_{0x}/\gamma$, etc. If $\gamma \ll 1$, then we have a slow–fast system with the repressor acting as the slow variable. The existence of a relaxation oscillator can then be established using phase-plane analysis.

6.4.4 The Circadian Clock and Molecular Noise

The circadian rhythm plays a key physiological role in the adaptation of living organisms to the alternation of night and day [214, 492]. Experimental studies of a wide range of plants and animals has established that in almost all cases, autoregulatory feedback on gene expression plays a central role in the molecular mechanisms

underlying circadian rhythms [335, 501]. Based on experimental data, a variety of models of increasing complexity have been developed, which show how regulatory feedback loops in circadian gene networks generate sustained oscillations under conditions of continuous darkness [188, 222, 381, 382, 602, 648]. The resulting circadian oscillator has a natural period of approximately 24 h, which can be entrained to the external light–dark cycle. Given that the circadian rhythm is controlled by gene networks, this immediately raises the issue regarding the extent to which such oscillations are robust to intrinsic noise arising from small numbers of molecules [21, 187, 189, 227]. Here we review the analysis of Gonze et al. [228], who considered the effects of molecular noise on a minimal model of the circadian clock in the fungus *Neurospora* [382]. A schematic diagram of the basic model is shown in Fig. 6.11. A clock gene X (*frq* in Neurospora, *per* in Drosophila) is transcribed to form mRNA (M), which exits the nucleus and is subsequently translated into cytoplasmic clock protein (X_C). The resulting protein either degrades or enters the nucleus (X_N) where it inhibits its own gene expression.

The governing equations for the concentrations m, x_C, x_N of mRNA, cytosolic protein, and nuclear protein, respectively, are

$$\frac{dm}{dt} = k\frac{K_m^n}{K_m^n + x_N^n} - \gamma\frac{m}{K_m' + m} \tag{6.4.16a}$$

$$\frac{dx_C}{dt} = rm - \gamma_P\frac{x_C}{K_p + x_C} - k_1 x_C + k_2 x_N \tag{6.4.16b}$$

$$\frac{dx_N}{dt} = k_1 x_C - k_2 x_N. \tag{6.4.16c}$$

Here k is the unregulated rate of transcription, r is the rate of translation, and γ, γ_p are the rates of mRNA and protein degradation; degradation is assumed to obey Michaelis–Menten kinetics. The negative regulation of transcription is taken to be cooperative with a Hill coefficient of n. Finally, the rate constants k_1, k_2 characterize the transport of protein into and out of the nucleus. It can be shown that the above model exhibits limit cycle oscillations in physiologically reasonable parameter regimes and thus provides a molecular basis for the sustained oscillations of the circadian clock under constant darkness [382]. In order to explore the robustness of such oscillations to molecular noise, it is necessary to turn to a master equation formulation of the gene network. One can then approximate the master equation by an FP equation as outlined in Box 6A, but now one has to linearize the FP equation about a limit cycle rather than a fixed point.

As in previous examples of gene regulation, it is convenient to rewrite this system of equations in the form of Eq. (6.3.17), which involves a sum over $R = 6$ single-step reactions labeled by a, whose transition rates $f_a(\mathbf{x})$ and stochiometric coefficients S_{ia} are listed in the table below with $\mathbf{x} = (m, x_C, x_N)^T$ [228]. Given this decomposition, one can now write down the FP equation obtained under the diffusion approximation [see also Eq. (6.3.20)]:

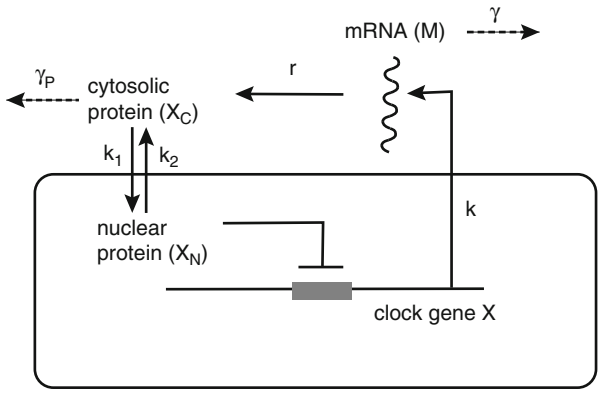


Fig. 6.11: Minimal model for a negative autoregulation network underlying circadian rhythms. Transcription of a clock gene (X) produces mRNA (M), which is transported outside the nucleus and then translated into cytosolic clock protein (X_C). The protein is either degraded or transported into the nucleus (X_N) where it exerts negative feedback on the gene expression

$$\frac{\partial p}{\partial t} = -\sum_{i=1}^{3} \frac{\partial V_i(\mathbf{x}) p(\mathbf{x},t)}{\partial x_i} + \frac{1}{2\Omega} \sum_{i,j=1}^{3} \frac{\partial^2 D_{ij}(\mathbf{x}) p(\mathbf{x},t)}{\partial x_i \partial x_j}, \tag{6.4.17}$$

where Ω is the total number of molecules that can be present in the system, say,

$$V_i(\mathbf{x}) = \sum_{a=1}^{R} S_{ia} f_a(\mathbf{x}), \quad D_{ij}(\mathbf{x}) = \sum_{a=1}^{R} S_{ia} S_{ja} f_a(\mathbf{x}). \tag{6.4.18}$$

From Table 6.1, we deduce that

$$V_1(\mathbf{x}) = f_1(\mathbf{x}) - f_2(\mathbf{x}), V_2(\mathbf{x}) = f_3(\mathbf{x}) - f_4(\mathbf{x}) - f_5(\mathbf{x}) + f_6(\mathbf{x}), V_3(\mathbf{x}) = f_5(\mathbf{x}) - f_6(\mathbf{x}),$$

and

$$\begin{aligned}
D_{11}(\mathbf{x}) &= \frac{1}{2}(f_1(\mathbf{x}) + f_2(\mathbf{x}), \quad D_{12} = D_{21} = D_{13} = D_{31} = 0, \\
D_{22}(\mathbf{x}) &= \frac{1}{2}(f_3(\mathbf{x}) + f_4(\mathbf{x}) + f_5(\mathbf{x}) + f_6(\mathbf{x})) \\
D_{23}(\mathbf{x}) &= D_{32}(\mathbf{x}) = -\frac{1}{2}(f_5(\mathbf{x}) + f_6(\mathbf{x})), \quad D_{33}(\mathbf{x}) = \frac{1}{2}(f_5(\mathbf{x}) + f_6(\mathbf{x})).
\end{aligned}$$

In the deterministic limit $\Omega \to \infty$, we recover the deterministic model (6.4.16), which can be rewritten in the more compact form

$$\frac{dx_i}{dt} = V_i(\mathbf{x}), \quad i = 1,2,3. \tag{6.4.19}$$

Reaction	Transition rate	Transition
$X \to X + M$	$f_1 = k\dfrac{K_m^n}{K_m^n + x_N^n}$	$M \to M + 1$
$M \to \emptyset$	$f_2 = \gamma\dfrac{m}{K_m' + m}$	$M \to M - 1$
$M \to X_C + M$	$f_3 = rm$	$X_C \to X_C + 1$
$X_C \to \emptyset$	$f_4 = \gamma_P\dfrac{x_C}{K_p + x_C}$	$X_C \to X_C - 1$
$X_C \to X_N$	$f_5 = k_1 x_C$	$X_C \to X_C - 1,\ X_N \to X_N + 1$
$X_N \to X_C$	$f_6 = k_2 x_N$	$X_C \to X_C + 1,\ X_N \to X_N - 1$

Table 6.1: Single-step reactions of the minimal circadian clock gene network

One way to investigate the effects of molecular noise on the circadian clock is to linearize the FP equation about the limit cycle solution, analogous to the linear noise approximation for Gaussian-like fluctuations about fixed points (see Box 6A). However, the linear noise approximation requires that perturbations remain small for all times, which is not the case for limit cycles, since they are marginally stable with respect to phase shifts around the limit cycle. Therefore, one needs to separate out the effects of longitudinal and transverse fluctuations of the limit cycle [56, 578]. The basic intuition is that Gaussian-like transverse fluctuations are distributed in a tube of radius $1/\sqrt{\Omega}$, whereas the phase around the limit cycle undergoes Brownian diffusion. Thus, consider the Langevin equation corresponding to the FP equation (6.4.17):

$$dX_i(t) = V_i(\mathbf{X}(t))dt + \frac{1}{\sqrt{\Omega}}\sum_{j=1}^{n} D_{ij}(\mathbf{X}(t))dW_j(t), \tag{6.4.20}$$

where n is the number of chemical species ($n = 3$ in the case of the circadian clock) and $W_j(t)$ are independent Wiener processes. We can then decompose the stochastic vector $\mathbf{X}(t)$ according to

$$\mathbf{X}(t) = \mathbf{x}^*(t + S(t)) + \mathbf{T}(t), \tag{6.4.21}$$

where the scalar random variable $S(t)$ represents the undamped random phase shift along the limit cycle and $\mathbf{T}(t)$ is a transversal perturbation (see Fig. 6.12). Since there is no damping of fluctuations along the limit cycle, the random phase $S(t)$ is taken to undergo Brownian motion. The associated phase diffusion coefficient D_θ is an effective time constant that characterizes the robustness of the oscillator to intrinsic noise. However, it is important to note that the decomposition (6.4.21) is not unique, so that the precise definition of the phase depends on the particular method of analysis.

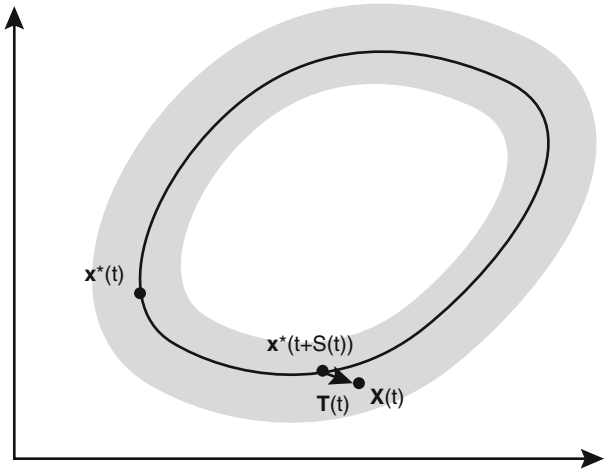


Fig. 6.12: Decomposition of a stochastic limit cycle $\mathbf{X}(t)$ into a random phase shift $S(t)$ along the deterministic limit cycle $\mathbf{x}^*(t)$ and a random transversal component $\mathbf{T}(t)$

For example, one recent study defines the phase in order to ensure that the mean size of transversal fluctuations remains small [343]. On the other hand, Gonze et al. [228] estimate D_θ for their circadian clock model using an alternative approach based on a WKB approximation of solutions to the FP equation (see also [210, 650]). In particular, they find $D_\theta \sim 1/\Omega$ so that larger systems are more robust to noise as one would expect. Gonze et al. also determine the rate of decay of correlations. Let $\mathbf{x}_0$ be a point on the deterministic limit cycle and suppose $\mathbf{X}(0) = \mathbf{x}_0$. Define the rth return time τ_r of a trajectory to be when an arbitrarily chosen component $X_j(t)$ returns to $x_{j,0}$ for the rth time, $X_j(\tau_r) = x_{j,0}$. In the deterministic case, $\tau_r = rT$ where T is the minimal period of the limit cycle oscillation. For sufficiently large Ω (small noise), we expect the distribution of first return times $\tau_1 = \tau$ to be approximately given by the Gaussian

$$P(\tau) \sim \frac{1}{\sqrt{2\pi D_\theta T}} \exp\left[-\frac{(\tau - T)^2}{2D_\theta T}\right].$$

It can also be shown that the autocorrelation function for each concentration takes the form of damped oscillations [210, 228],

$$\begin{aligned} C_j(t) &= \langle X_j(t) X_j(0) \rangle \\ &\approx C_{j,0} + C_{j,1} \mathrm{e}^{-t/\gamma} \cos(\omega t + \alpha_j), \quad t \to \infty, \end{aligned} \tag{6.4.22}$$

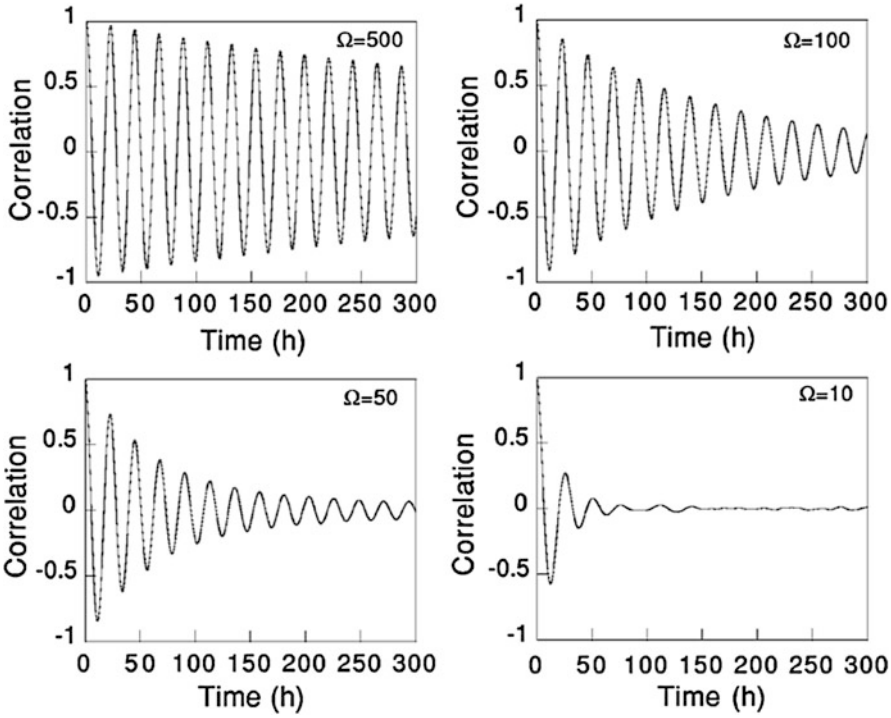


Fig. 6.13: Illustration of the time evolution of one of the autocorrelation functions of the stochastic circadian clock model considered by Gonze et al. [228]. As the system size Ω is decreased, the rate of decay of correlations becomes more rapid. Parameter values can be found in [228]

for some coefficients $C_{j,0}, C_{j,1}$ and phases α_j, with

$$\omega \approx \frac{2\pi}{T}, \quad \gamma \approx \frac{T^2}{2D_\theta \pi^2}. \tag{6.4.23}$$

It follows that the rate of decay γ^{-1} of correlations is inversely proportional to the system size (see Fig. 6.13).

6.4.5 Quasi-Cycles in a Biochemical Oscillator

In Sect. 6.4.4, we discussed the effects of intrinsic noise on a biochemical limit cycle oscillator that exists in the absence of noise. One also finds that stochastic biochemical and gene networks can exhibit noise-induced oscillations (quasi-cycles) in parameter regimes for which the underlying deterministic kinetic equations have only fixed point solutions. These quasi-cycles are characterized by a peak in the

power spectrum obtained using a linear noise approximation of the chemical master equation. Following Boland et al. [55, 56], we will illustrate this using a stochastic version of the Brusselator. (A spatially extended version of the model will be considered in Sect. 9.3.) The Brusselator is an idealized model of an autocatalytic reaction, in which at least one of the reactants is also a product of the reaction [220]. The model consists of two chemical species X and Y interacting through the following reaction scheme:

$$\begin{aligned} \emptyset &\xrightarrow{a} X \\ X &\xrightarrow{b} Y \\ 2X + Y &\xrightarrow{c} 3X \\ X &\xrightarrow{d} \emptyset \end{aligned}$$

These reactions describe the production and degradation of an X molecule, an X molecule spontaneously transforming into a Y molecule, and two molecules of X reacting with a single molecule of Y to produce three molecules of X. The corresponding mass-action kinetic equations for $u = [X], v = [Y]$ are (after rescaling so that $c = d = 1$)

$$\frac{du}{dt} = a - (b+1)u + u^2 v, \tag{6.4.24a}$$

$$\frac{dv}{dt} = bu - u^2 v. \tag{6.4.24b}$$

The system has a fixed point at $u^* = a, v^* = b/a$, which is stable when $b < a^2 + 1$ and unstable when $b > a^2 + 1$ (see below). Moreover, the fixed point undergoes a Hopf bifurcation at the critical value $b = a^2 + 1$ for fixed a, leading to the formation of a stable limit cycle (see Box 4B).

Following the examples of Sect. 6.3, it is straightforward to write down a stochastic version of the Brusselator model. Let $n_1(t)$ and $n_2(t)$ denote the number of X and Y molecules at time t, respectively, and take Ω to be cell volume. The various state transitions are

$$(n_1, n_2) \xrightarrow{T_1} (n_1 + 1, n_2), \tag{6.4.25a}$$

$$(n_1, n_2) \xrightarrow{T_2} (n_1 - 1, n_2 + 1) \tag{6.4.25b}$$

$$(n_1, n_2) \xrightarrow{T_3} (n_1 + 1, n_2 - 1) \tag{6.4.25c}$$

$$(n_1, n_2) \xrightarrow{T_4} (n_1 - 1, n_2), \tag{6.4.25d}$$

with $\mathbf{n} = (n_1, n_2)$ and

$$T_1 = \Omega a, \quad T_2 = b n_1, \quad T_3 = n_1^2 n_2 / \Omega^2, \quad T_4 = n_1. \tag{6.4.26}$$

It is convenient to rewrite the kinetic equations (6.4.24) in the generic form (6.3.17):

$$\frac{du_i}{dt} = \sum_{r=1}^{p} S_{ir} f_r(u_1, u_2), \quad i = 1,2 \tag{6.4.27}$$

where $u_j = n_j/\Omega$, r labels the single-step reaction, p is the number of single-step reactions, and $\mathbf{S}$ is the stochiometric matrix. In the case of the Brusselator model, Eq. (6.4.24) shows that $p = 4$,

$$S_{11} = 1, S_{21} = 0, \quad S_{12} = -1, S_{22} = 1, \quad S_{13} = 1, S_{23} = -1, \quad S_{14} = -1, S_{24} = 0,$$

and $\Omega f_r(n_1/\Omega, n_2/\Omega) = T_r(n_1, n_2)$. The corresponding master equation is then given by [see Eq. (6.3.18)],

$$\frac{dP(n_1,n_2,t)}{dt} = \Omega \sum_{r=1}^{p} \left(\prod_{i=1,2} \mathbb{E}^{-S_{ir}} - 1 \right) f_r(n_1/\Omega, n_2/\Omega) P(n_1, n_2, t). \tag{6.4.28}$$

Now suppose that Ω is sufficiently large so that we can carry out a linear noise approximation and obtain a Langevin equation for a multivariate OU process (see Box 6A). That is, we approximate the master equation (6.4.28) by an FP equation and then linearize about the fixed point (u^*, v^*) by setting

$$\frac{n_j}{\Omega} = u_j = u_j^* + \frac{1}{\sqrt{\Omega}} v_j.$$

This yields the Langevin equation

$$\frac{dv_j(t)}{dt} = \sum_{j'} A_{jj'} v_{j'}(t) + \eta_j(t), \tag{6.4.29}$$

with white noise terms satisfying

$$\langle \eta_j(t) \rangle = 0, \quad \langle \eta_j(t) \eta_{j'}(t') \rangle = D_{jj'} \delta(t - t').$$

Here

$$A_{jj'} = \sum_{r=1}^{p} S_{jr} \frac{\partial f_r(u_1^*, u_2^*)}{\partial u_{j'}^*} \tag{6.4.30}$$

and

$$D_{jj'} = \sum_{r=1}^{p} S_{jr} S_{j'r} f_r(u_1^*, u_2^*). \tag{6.4.31}$$

Fourier transforming the Langevin equation with respect to time using

$$V_i(\omega) = \int_{-\infty}^{\infty} \mathrm{e}^{i\omega t} v_i(t) dt$$

etc. gives

$$\sum_l \Phi_{jl}(\omega)V_l(\omega) = \eta_j(\omega)$$

with

$$\Phi_{jl}(\omega) = -i\omega\delta_{j,l} - A_{jl}$$

and

$$\langle \eta_j(\omega)\rangle = 0, \quad \langle \eta_j(\omega)\eta_{j'}(\omega')\rangle = D_{jj'}\delta(\omega+\omega').$$

Hence,

$$\begin{aligned}\langle V_i(\omega)V_i(\omega')\rangle &= \left\langle \left[\sum_l \Phi_{il}^{-1}(\omega)\eta_l(\omega)\right]\left[\sum_j \Phi_{ij}^{-1}(\omega')\eta_j(\omega')\right]\right\rangle \\ &= \delta(\omega+\omega')\sum_{l,j}\Phi_{il}^{-1}(\omega)D_{lj}\Phi_{ij}^{-1}(-\omega').\end{aligned}$$

Defining the power spectrum of the ith chemical species by (see Sects. 2.2.4 and 6.3)

$$\langle V_i(\omega)V_i(\omega')\rangle = S_i(\omega)\delta(\omega+\omega'),$$

we deduce that

$$S_i(\omega) = \sum_{l,j}\Phi_{il}^{-1}(\omega)D_{lj}(\Phi^{\dagger})_{ji}^{-1}(\omega). \tag{6.4.32}$$

Note that the above analysis applies to any two-species master equation of the form (6.4.28) and can be extended to multiple species. In the case of the Brusselator model system, whose deterministic kinetic equations are given by Eq. (6.4.24), we have

$$\sum_r S_{1r}f_r(u_1,u_2) = a-(b+1)u_1+u_1^2u_2, \tag{6.4.33}$$

$$\sum_r S_{2r}f_r(u_1,u_2) = bu_1-u_1^2u_2. \tag{6.4.34}$$

It follows that

$$\mathbf{A} = \begin{pmatrix} b-1 & a^2 \\ -b & -a^2 \end{pmatrix}, \tag{6.4.35}$$

and

$$\mathbf{D} = \begin{pmatrix} 2(b+1)a & -2ba \\ -2ba & 2ba \end{pmatrix} \tag{6.4.36}$$

on setting $u_1^* = a, u_2^* = b/a$. The corresponding power spectra are [55]

$$S_1(\omega) = 2a((1+b)\omega^2+a^4)\Gamma(\omega)^{-1}, \quad S_2(\omega) = 2ab(\omega^2+1+b)\Gamma(\omega)^{-1}, \tag{6.4.37}$$

where

$$\Gamma(\omega) = (a^2-\omega^2)^2+(1+a^2-b)^2\omega^2.$$

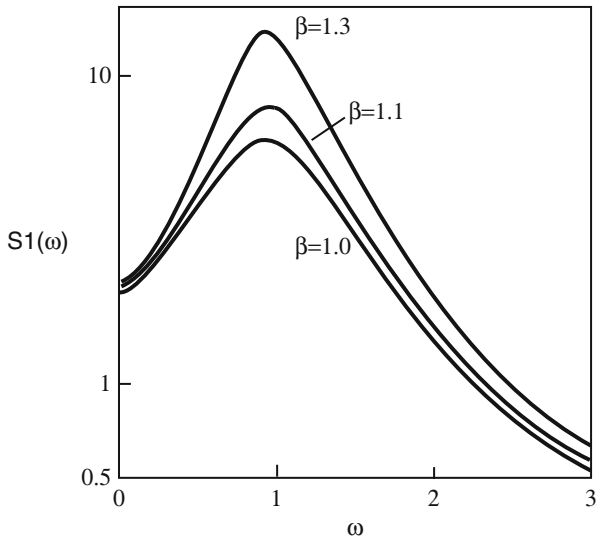


Fig. 6.14: Power spectrum $S_1(\omega)$ of fluctuations in the concentration of X molecules in Brusselator system for parameter values in the fixed point regime of the kinetic equations (6.4.24): $a = 1$ and $b = 1.0, 1.1, 1.3$

In Fig. 6.14 we plot the power spectrum $S_1(\omega)$ in a parameter regime where the deterministic kinetic equation (6.4.24) support a stable fixed point. It can be seen that there is a peak in the power spectrum at $\omega = \omega_c \neq 0$, indicating the presence of stochastic oscillations (quasi-cycles) even though the deterministic system operates below the Hopf bifurcation point. Moreover the frequency ω_c is approximately equal to the Hopf frequency of limit cycle oscillations beyond the bifurcation point. Thus, intrinsic noise can extend the parameter regime over which a biochemical system can exhibit oscillatory behavior. An analogous result holds for Turing pattern formation, as discussed in Sect. 9.3.

6.5 Information Transmission in Regulatory Networks

So far in this chapter we have focused on methods for calculating the level of molecular noise in gene networks. However, as we indicated in the introduction, one important consequence of noise is that it can limit the ability of a network to transmit information. In this section we show how to mathematically quantify this idea, following closely the recent review by Tkacik and Walczak [631]. In order to motivate the analysis, let us return to the simple feedforward regulatory network considered in Sect. 6.3.2 and Fig. 6.3. We now make explicit the fact that the rate at which the gene switches to its active state will depend on the background concentration $c(t)$ of the transcription factor Y. Therefore, we set $k_+ \to k_+ c$ and treat the network

as an input/output device $c(t) \to x(t)$, where $x(t)$ is the protein concentration (see Fig. 6.15). In the case of promoter switching that is faster than the protein kinetics and time variation of c, the probability p_1 that the gene is active reaches the QSS $p_1(c) = k_+c/(k_+c+k_-)$. Therefore, ignoring intrinsic fluctuations, the protein concentration evolves according to the kinetic equation

$$\frac{dx}{dt} = rk_+c(t)/(k_+c(t)+k_-) - \gamma x(t). \tag{6.5.1}$$

In the case of a constant input, the steady-state solution is

$$\bar{x}(c) \equiv \frac{r}{\gamma}\frac{k_+c}{k_+c+k_-}, \tag{6.5.2}$$

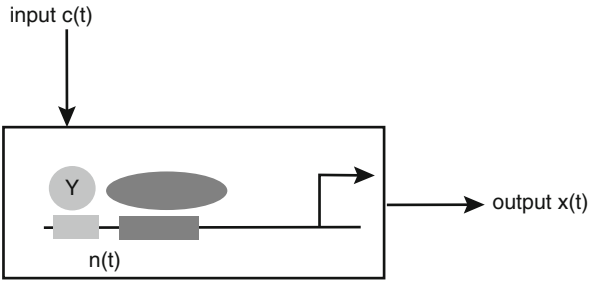


Fig. 6.15: Simple regulatory network represented as a noisy input/output channel. The input signal is the concentration $c(t)$ of transcription factor and the output signal is the concentration $x(t)$ of expressed protein. The internal state $n(t)$ of the channel specifies whether the gene is in the active ($n = 1$) or inactive ($n = 0$) state

which is an invertible function. Hence, we can faithfully reconstruct the input from the output—we have lossless channel. However, this no longer holds when the effects of intrinsic noise are included. Given a fixed input and weak noise, the output response can be characterized in terms of the stationary Gaussian distribution

$$P(x|c) = \frac{1}{\sqrt{2\pi\sigma_X^2(c)}} \mathrm{e}^{-(x-\bar{x}(c))^2/2\sigma_X^2(c)}, \tag{6.5.3}$$

where $\sigma_X^2(c)$ is the variance of the protein concentration. In Sect. 6.3.2 we calculated the variance in the case of a large but finite population of genes using a linear noise approximation. We found that there were two contributions to the variance—input noise arising from the stochastic binding and unbinding of transcription factors and output Poisson noise due to the finite number of proteins [see Eq. (6.3.16)]. Both terms depend on c under the substitution $k_+ \to k_+c$. It turns out that analogous terms arise in the case of a single gene. However, the input noise tends to be smaller

than diffusion noise, which arises from the passive transport of transcription factors within the nucleus [632]. Therefore, applying the Berg–Purcell theory of diffusion-limited reactions to estimate the size of fluctuations in the input concentration c (see Sect. 5.1), we have $\sigma_c^2 = c/(Da\tau)$ where D is the diffusivity, a is the size of the promoter binding site, and τ is the detection time (taken to be the lifetime of a protein). The total output variance is then

$$\sigma_X^2(c) = \bar{x}(c) + \left(\frac{d\bar{x}(c)}{dc}\right)^2 \sigma_c^2, \tag{6.5.4}$$

where the factor $(d\bar{x}(c)/dc)^2$ converts input fluctuations to output fluctuations in the small noise limit.

Suppose that the inputs are drawn from some stationary distribution $P(c)$. A fundamental issue is finding a way to quantify how much information one can extract, in principle, about the values of the input c based on measurements of the output x, given the conditional probability $P(x|c)$. Since the joint probability distribution is $P(x,c) = P(x|c)P(c)$, it follows that if x and c were statistically independent (c-independent $\bar{x}$ and σ_X^2), then $P(c,x) = P(c)P(x)$ and $P(x|c) = P(x)$. In this case the channel cannot transmit any information. This suggests that quantifying the amount of information transmitted involves some measure of the statistical interdependence of the inputs and outputs. A first guess might be the covariance

$$\mathrm{Cov}(c,x) = \int\int (c - \langle c\rangle)(x - \langle x\rangle) dx\, dc.$$

However, this only captures linear correlations, whereas many gene networks are nonlinear input/output devices. A much more general measure of statistical interdependence, which is based on a minimal number of assumptions about the underlying stochastic process, was introduced by Shannon [582] and is known as the *mutual information* between c and x.

6.5.1 Entropy and Mutual Information

In Sect. 1.4 we introduced the notion of entropy in terms of the number of different configurations M that a macromolecule with given energy E can realize, that is, $S = k_B \ln M$. (In statistical mechanics this is referred to as the entropy of a microcanonical ensemble.) An implicit assumption is that each of these microstates i is equally likely so the probability distribution over the set of microstates for fixed E is uniform, $p_i = 1/M$. Hence, we could rewrite the entropy as

$$S = -k_B \sum_i p_i \ln p_i. \tag{6.5.5}$$

It turns out that such a formula also applies to systems with different constraints than fixed energy, where the distribution p_i is not uniform. One example is the Boltzmann–Gibbs distribution itself, where one allows the energy to fluctuate but the mean energy is fixed by the temperature of the surrounding environment. The basic idea is that the entropy still counts the number of accessible states but weights them according to the probability of observing a given state. Shannon subsequently introduced an information theoretic notion of entropy as a measure of the uncertainty of a random variable; the larger the entropy, the greater the amount of information that is generated by observing the state of the system. The convention in information theory is to set $k_B = 1$ and to use base 2 logarithms. Thus, the *Shannon entropy* is defined according to

$$S = -\sum_i p_i \log_2 p_i \tag{6.5.6}$$

so that S is measured in bits. One bit is the entropy of a binary variable that has two equally accessible states. In the case of M possible states, the entropy takes values in the range $0 \leq S \leq \log_2 M$. If $S = 0$, then there is no uncertainty and making a measurement yields no new information. On the other hand, the entropy is maximal when $p_i = 1/M$. It is also possible to define Shannon entropy for a continuous random variable such as the concentration c of a protein:

$$S = -\int p(c) \log_2 p(c) dc. \tag{6.5.7}$$

(Note that certain care has to be taken, since the entropy depends on the units chosen for the continuous random variable. However, we will be interested in changes in entropy where this is no longer an issue.)

Recall that we want to find some measure of the statistical interdependence of an input c and an output x. From the perspective of information theory, one can quantify this in terms of how much one's uncertainty in x is reduced by knowing c. Prior to knowing c, the entropy is $S[P_X] = -\int P(x) \log_2 P(x) dx$, whereas after c is specified the entropy becomes $S[P_{X|C}] = -\int P(x|c) \log_2 P(x|c) dx$. Thus, a measure of the reduction in uncertainty is the entropy difference

$$\Delta S = S[P_X] - S[P_{X|C}].$$

Now imagine measuring the entropy difference over an ensemble of different input concentration regimes distributed according to $P(c)$. The resulting quantity is called the *mutual information*

$$I(c;x) = \int P(c)(S[P_X] - S[P_{X|C}]) dc. \tag{6.5.8}$$

One important property of the mutual information is that it is symmetric with respect to exchanging input and output. In order to see this, we use the definition of Shannon entropy, and the relation between joint and conditional probabilities:

$$
\begin{aligned}
I(c;x) &= -\int\int P(c)\,[P(x)\log_2 P(x) - P(x|c)\log_2 P(x|c)]\,dx\,dc \\
&= \int\int \left[P(c,x)\log_2 \frac{P(c,x)}{P(c)} - P(c,x)\log_2 P(x)\right] dx\,dc \\
&= \int\int \left[P(c,x)\log_2 \frac{P(c,x)}{P(c)P(x)}\right] dx\,dc = \int\int \left[P(c,x)\log_2 \frac{P(c|x)}{P(c)}\right] dx\,dc \\
&= \int\int [P(c|x)P(x)\log_2 P(c|x) - P(c,x)\log_2 P(c)]\,dx\,dc \\
&= -\int\int P(x)\,[P(c)\log_2 P(c) - P(c|x)\log_2 P(c|x)]\,dc\,dx.
\end{aligned}
$$

Hence, the mutual information measures how much, on average, our uncertainty in one variable is reduced by knowing the value of the complementary variable.

Example 6.1 (Gaussian noise). As an illustration of the above ideas, suppose that an input signal is corrupted by additive Gaussian noise

$$x = \gamma c + \xi.$$

This means that the conditional probability distribution is

$$P(x|c) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\gamma c)^2}{2\sigma^2}\right).$$

Also suppose that the input c is also drawn from a Gaussian distribution,

$$P(c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left(-\frac{c^2}{2\sigma_c^2}\right).$$

It follows that x is also a Gaussian with

$$P(x) = \int P(x|c)P(c)dc = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{x^2}{2\sigma_x^2}\right)$$

with

$$\sigma_x^2 = \gamma^2\sigma_c^2 + \sigma^2.$$

The mutual information is

$$
\begin{aligned}
I(c;x) &= \frac{1}{\ln 2}\int\int \left[P(c,x)\ln\frac{P(x|c)}{P(x)}\right] dx\,dc \\
&= \frac{1}{\ln 2}\int\int P(c,x)\left[\ln\left(\sqrt{\frac{2\pi\sigma_x^2}{2\pi\sigma^2}}\right) - \frac{(x-\gamma c)^2}{2\sigma^2} + \frac{x^2}{2\sigma_x^2}\right] dx\,dc,
\end{aligned}
$$

on using the identity $\log_2 x = \ln x / \ln 2$. Using the results $\int\int P(c,x)x^2 dx\,dc = \langle x^2 \rangle$, $\int\int P(x,c)dx\,dc = 1$, and

$$\int\int P(c,x)(x-\gamma c)^2 dx\,dc = \int\int P(x|c)P(c)(x-\gamma c)^2 dx\,dc = \sigma^2,$$

we deduce that

$$I(c;x) = \frac{1}{\ln 2} \ln\left(\sqrt{\frac{\sigma_x^2}{\sigma^2}}\right) = \frac{1}{2}\log_2\left(1+\frac{\gamma^2\sigma_c^2}{\sigma^2}\right). \tag{6.5.9}$$

The mutual information depends on the so-called SNR σ_c^2/σ^2. It can also be shown that in the case of Gaussian additive noise, information transmission or mutual information is maximized for a given input variance when the input is drawn from a Gaussian distribution [631].

6.5.2 *Optimizing Mutual Information in a Simple Regulatory Network*

A basic goal of information theory within the context of gene regulatory networks is to determine the distribution of inputs for which a network with a given form of intrinsic noise maximizes its information transmission as measured by mutual information. This program has been developed for a range of networks with ever increasing complexity, including multiple gene products that may interact, and self-regulatory feedback [632–635, 665]. For related studies on information transmission, see [105, 383, 638, 705]. In order to make analytical progress, the system is usually assumed to operate in the small noise regime so that the various conditional probability distributions can be approximated by Gaussians. Here we will illustrate the basic ideas by focusing on the simple regulatory network of Fig. 6.15, following the analysis of Tkacik et al. [634]. Since we are optimizing with respect to $P(c)$, we use the following version of the mutual information:

$$I(x;c) = -\int P(c)\log_2 P(c)dc + \int\int P(c|x)\log_2 P(c|x)dc\,dx. \tag{6.5.10}$$

However, to use this formula, it is necessary to determine $P(c|x)$ given $P(x,c)$. Exploiting the small noise approximation, we model $P(c|x)$ as the Gaussian

$$P(c|x) = \frac{1}{\sqrt{2\pi\sigma_C^2(x)}} e^{-(c-\bar{c}(x))^2/2\sigma_C^2(x)}, \tag{6.5.11}$$

where $\bar{c}(x)$ is the most likely value of c given the output x and $\sigma_C^2(x)$ is the corresponding variance about the expected value. Substituting into the expression for the mutual information yields

$$I(x;c) = -\int P(c)\log_2 P(c)dc - \frac{1}{2}\int P(x)\log_2[2\pi e\sigma_C^2(x)]dx.$$

It remains to determine $\bar{c}(x)$ and $\sigma_C^2(x)$. Using Bayes' theorem (Sect. 1.3), we have

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} = \frac{1}{Z(x)}\mathrm{e}^{-F(c,x)},$$

where all c-independent terms have been lumped together in the normalization factor Z and, from Eq. (6.5.3),

$$F(c,x) = -\ln P(c) + \frac{1}{2}\ln\sigma_X^2(c) + \frac{1}{2}\frac{(x-\bar{x}(c))^2}{\sigma_X^2(c)}.$$

Comparison with Eq. (6.5.11) shows that $\bar{c}(x)$ and $\sigma_C^2(x)$ are defined by

$$\left.\frac{\partial F(c,x)}{\partial c}\right|_{c=\bar{c}(x)=0} = 0, \quad \left.\frac{\partial^2 F(c,x)}{\partial c^2}\right|_{c=\bar{c}(x)=0} = \frac{1}{\sigma_C^2(x)}.$$

Expanding the solution for $1/\sigma_C^2(x)$ to leading order in $1/\sigma_X^2(c)$ (small noise approximation) then gives [634]

$$\frac{1}{\sigma_C^2(x)} = \frac{1}{\Sigma^2(c)} \equiv \frac{1}{\sigma_X^2(c)}\left(\frac{d\bar{x}(c)}{dc}\right)^2 \tag{6.5.12}$$

and hence

$$I(x;c) = -\int P(c)\log_2 P(c)dc + \frac{1}{2}\int P(c)\log_2\left[\frac{1}{2\pi e}\frac{1}{\Sigma^2(c)}\right]dc. \tag{6.5.13}$$

We have used the fact that, to leading order, $P(x)dx = P(c)dc$ in the small noise limit.

We now have the variational problem of finding the input distribution $P^*(c)$ that maximizes the mutual information. However, it is first necessary to specify certain constraints regarding the optimization procedure. First, there is a maximum number of proteins involved, which can be imposed by restricting the allowed range of the input concentration c and the corresponding expected number of proteins $\bar{x}(c)$; the latter is typically implemented by normalizing $\bar{g}(c)$ to lie in the interval $[0,1]$. The constraint that $\int_0^{c_{\max}} P(c)dc = 1$ can be incorporated into the variational problem using a Lagrange multipler λ, so that the optimization problem takes the form

$$\frac{\delta}{\delta P(c)}\left[I(x;c) - \lambda\int P(c)dc\right] = 0. \tag{6.5.14}$$

Using the expression for $I(x;c)$ and properties of functional derivatives[1], one finds that the optimal input distribution $P^*(c)$ satisfies

$$0 = \frac{1}{2}\ln\left[\frac{1}{2\pi e}\frac{1}{\Sigma^2(c)}\right] - \ln P^*(c) - 1 - \lambda \ln 2.$$

Rearranging and exponentiating gives the optimal distribution

$$P^*(c) = \frac{1}{Z}\frac{1}{\sqrt{2\pi e}}\frac{1}{\Sigma(c)}, \tag{6.5.15}$$

where $Z = e^{1+\lambda \ln 2}$ is a normalization factor with

$$Z = \int_0^{c_{\max}} \frac{1}{\sqrt{2\pi e}}\frac{dc}{\Sigma(c)}. \tag{6.5.16}$$

The corresponding optimal mutual information is simply

$$I^* = \log_2 Z. \tag{6.5.17}$$

Note that the resulting expression for I^* will still depend on the various parameters of the underlying regulatory network. These include the parameters associated with the kinetics of binding and unbinding of transcription factors as in Eq. (6.5.4) and, in more complex networks, interactions between multiple gene products. Thus, there is an additional optimization step in which one maximizes the mutual information I^* with respect to these network parameters. Some of the predictions regarding the structure of optimal networks can be found elsewhere [105, 634, 635, 665, 705]. Here we illustrate the theory with the simple regulatory network given by Fig. 6.15. From Eqs. (6.5.4) and (6.5.16),

$$Z = \int_0^{c_{\max}} \frac{1}{\sqrt{2\pi e}}\frac{(d\bar{x}(c)/dc)^2}{\bar{x}(c) + c\,(d\bar{x}(c)/dc)^2}. \tag{6.5.18}$$

[1] The mutual information is expressed as a functional of $P(c)$, that is, $I = I[P]$. By analogy with the least-action principle of classical mechanics (see Chap. 10), we can take the functional derivative according to

$$\delta I = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}(I[P+\varepsilon\delta P] - I[P]),$$

where $\delta P(c)$ is an arbitrary smooth function. Evaluating the first term in I, after converting to natural logarithms, we have

$$\int [P(c)+\varepsilon\delta P(c)]\ln(P(c)+\varepsilon\delta P(c))dc - \int P(c)\ln P(c)dc = \varepsilon\int \delta P(c)[\ln P(c)+1]dc + O(\varepsilon^2).$$

Combining this with the other terms and using the fact that $\delta P(c)$ is arbitrary, we can set the total factor multiplying $\delta P(c)$ to be zero, which yields Eq. (6.5.15).

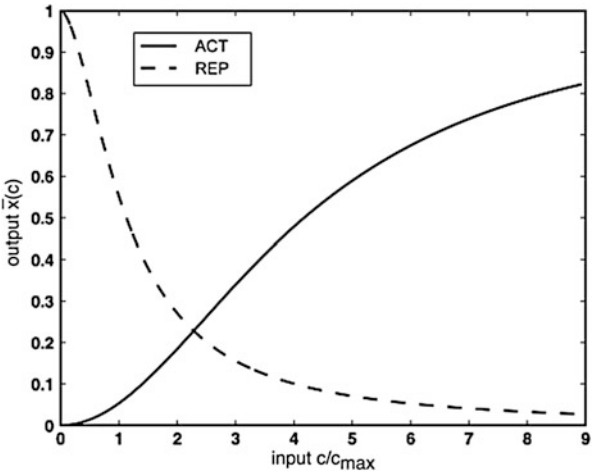


Fig. 6.16: The optimal input/output relations for repressors (*dashed line*) and activators (*solid line*) for the simple regulatory network shown in Fig. 6.15 (Adapted from Tkacik and Walczak [631])

Here c has been nondimensionalized by fixing $Da\tau$ and normalizing $\bar{x}$. Suppose that $\bar{x}(c)$ is given by the Hill function

$$\bar{x}(c) = \frac{c^n}{K^n + c^n}.$$

For fixed $c_{\max}$ the only free parameters are K and n. Optimizing the mutual information with respect to these parameters leads to the results shown in Fig. 6.16.

We end this brief detour into the world of information theory by briefly noting some potential limitations of the above approach, as highlighted in the review by Tkacik and Walczak [631]. First, there is no a priori reason why information should be identified as an appropriate measure of biological function. Second, even if it is an appropriate measure, it is possible that gene networks and biochemical signaling pathways have not yet become optimized for biological function through evolution. That is, the networks observed today simply reflect their particular evolutionary history rather than some history-independent optimization scheme. On the other hand, recent experiments concerning specific gene networks active during early development suggest that at least these networks operate close to the limits imposed by intrinsic noise [237]. Finally, in order to strengthen the links between theory and experiment, it will be necessary to confront the following theoretical challenges: (i) dealing with information transmission in time-dependent nonlinear networks; (ii) understanding information transmission in spatially inhomogeneous systems; (iii) extending analytical methods beyond the small noise limit; (iv) linking information transmission to other important measures of network function such as metabolic cost.

6.6 Fluctuations in DNA Transcription

One of the simplifications in the models of gene expression discussed so far is that the multistage processes underlying transcription and translation have been reduced to single-step processes with exponential waiting times (Poisson approximation). However, transcription (and also translation) can be broken up into at least three stages called initiation, elongation, and termination [128, 238]. During the initiation stage, RNAP binds to a promoter site on the DNA and unzips the double helix so that the strand of DNA to be transcribed is made accessible. Following the transcription of the first few nucleotides, the so-called transcription elongation complex (TEC) is formed, which consists of the RNAP, the DNA, and the emerging mRNA. This signals the beginning of the elongation phase where the TEC slides along the DNA, extending the transcript one nucleotide at a time. The process is terminated when a specific site is reached, for example, and the nascent mRNA is released. An implicit assumption of the single-step Poisson approximation of gene transcription is that the rate-limiting step is initiation. However, there is growing evidence from single-molecule experiments that initiation can be much faster than elongation [18]. Moreover, in vitro studies of *E. coli* RNAP have established that processive mRNA synthesis is often disrupted by transcriptional pauses that can last anything from 1s to more than 20s [258, 469]. In some cases, the pauses are linked with the reverse translocation of the RNAP along the DNA, a process known as *backtracking* [482]. These observations suggest that the distribution of transcription times might be non-exponential with heavy tails arising from the long transcriptional pauses.

Recently, there have been a number of stochastic models of the elongation stage and backtracking [413, 535, 552, 661, 690]. For the sake of illustration, we will focus on the model of Voliotis et al. [661], which is based on a master equation description of the dual processes of the TEC translocating along the DNA and the extension of the nascent mRNA via polymerization. A schematic diagram of the basic kinetics is shown in Fig. 6.17 in the simpler case that backtracking cannot occur. Suppose that n nucleotides have been transcribed and the active site of RNAP is at the end of the precursor mRNA chain. Denote this so-called pretranslocated state of the active site by $m = 0$. The active site can then shift one step beyond the precursor mRNA to form a posttranslocated state denoted by $m = 1$. It is now in a position to add the next nucleotide to the precursor mRNA by polymerization so that $n \to n+1$ and m resets to 0. The rates of polymerization and depolymerization are given by k_f and k_b, while the rates of forward and backward translocation are given by a and b. Let $P_{n,m}(t)$ be the probability of finding the TEC in state (n,m) at time t. The corresponding master equation is given by [661]

$$\frac{dP_{n,0}}{dt} = k_f P_{n-1,1} + bP_{n,1} - (k_b + a)P_{n,0} \tag{6.6.1a}$$

$$\frac{dP_{n,1}}{dt} = k_b P_{n+1,0} + aP_{n,0} - (k_f + b)P_{n,1}, \tag{6.6.1b}$$

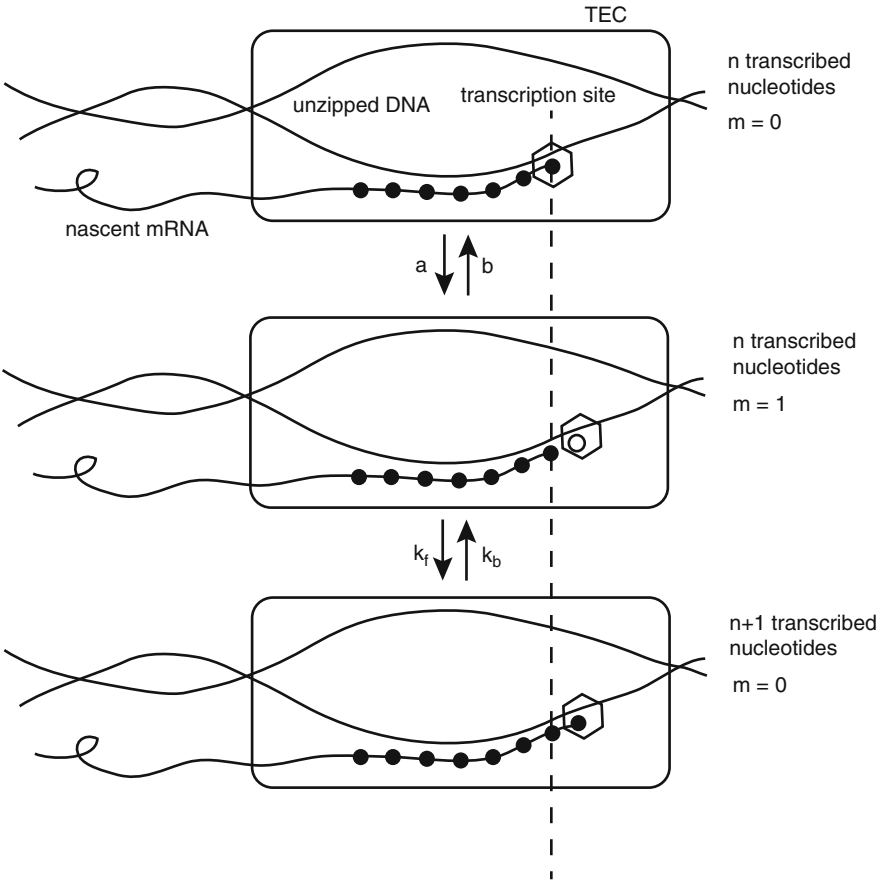


Fig. 6.17: Schematic illustration of the transcription elongation complex (*TEC*). In the absence of fluctuations, the active transcription site in the pretranslocated state ($m = 0$) takes one step beyond the nascent mRNA to enter the posttranslocated state ($m = 1$). Translocation is a reversible reaction with transition rates a and b. The site can then add one nucleotide to the mRNA via polymerization so that $n \to n+1$ and $m = 1 \to m = 0$. This step is also reversible with forward and backward transition rates k_f, k_b, respectively (Redrawn from [661])

with $n = 0, 1, \ldots N-1$. There is a reflecting boundary condition at $n = 0$, which can be implemented by introducing a fictitious state $n = -1$ and setting $k_b P_{0,0} = k_f P_{-1,1}$. Similarly, there is an absorbing boundary condition $P_{N,0} = 0$, since the process terminates when $n = N$ is reached.

The first step in the analysis is to introduce the mean occupancy for each translocation step ($m = 0, 1$) by summing over all nucleitide positions $n = 0, \ldots, N-1$. Setting $\Pi_m(t) = \sum_{n=0}^{N-1} P_{n,m}(t)$ and using the boundary conditions, we have

$$\frac{d\Pi_0}{dt} = (k_f + b)\Pi_1 - (k_b + a)\Pi_0, \quad \Pi_1 = 1 - \Pi_0$$

with the initial condition $\Pi_0(0) = 1$. There is convergence to the steady-state solution

$$\Pi_0^* = (k_f + b)\tau, \quad \Pi_1^* = (k_b + a)\tau, \quad \tau = \frac{1}{k_f + k_b + a + b},$$

with τ as the relaxation time. Assuming polymerization/depoylmerization is much slower than translocation ($k_f, k_b \ll a, b$), we can make the QSS approximation $P_{m,n}(t) = \Pi_m^* P_n(t)$ (see also Sect. 7.4), with $P_n(t)$ satisfying the birth–death master equation

$$\frac{dP_n}{dt} = \omega_- P_{n+1} + \omega_+ P_{n-1} - (\omega_+ + \omega_-)P_n, \tag{6.6.2}$$

and the effective polymerization/depolymerization rates are

$$\omega_+ = k_f(k_b + a)\tau \approx \frac{k_f a}{a+b}, \quad \omega_- = k_b(k_f + b)\tau \approx \frac{k_b b}{a+b}.$$

The boundary conditions become $\omega_- P_0 = \omega_+ P_{-1}$ (reflecting) and $P_N = 0$ (absorbing). The elongation time is defined as the time for the TEC to reach position $n = N$ starting from $n = 0$. In terms of the mean-field model given by the birth–death process, calculating the mean and variance of the elongation time requires solving a FPT for the discrete Markov process. This can be achieved by following analogous steps to the analysis of continuous process in Sect. 2.3.

Suppose that the TEC starts at position $n(0) = n_0$. Define the survival probability that the TEC has not yet reached the absorbing boundary at $n = N$ by

$$S(n_0, t) = \sum_{n=0}^{N-1} P(n, t | n_0, 0), \tag{6.6.3}$$

where we have made the initial condition explicit by setting $P_n(t) \to P(n, t | n_0, 0)$. If T is the (stochastic) elongation time, then $S(n_0, t)$ is the probability that $T \geq t$. This implies that the cumulative distribution function of the elongation time is $1 - S(n_0, 0)$. Hence the first and second moments of the elongation time are

$$T(n_0) = \langle T \rangle = \int_0^\infty t \frac{\partial S(n_0, t)}{\partial t} dt = \int_0^\infty S(n_0, t) dt \tag{6.6.4}$$

and

$$T_2(n_0) = \langle T^2 \rangle = \int_0^\infty t^2 \frac{\partial S(n_0, t)}{\partial t} dt = 2 \int_0^\infty t S(n_0, t) dt. \tag{6.6.5}$$

Equations for S and the moments of T can be obtained by considering the backward master equation

$$\begin{aligned} \frac{dP(n, t | n_0, 0)}{dt} &= \omega_+ \left[P(n, t | n_0 + 1, 0) - P(n, t | n_0, 0) \right] \\ &\quad + \omega_- \left[P(n, t | n_0 - 1, 0) - P(n, t | n_0, 0) \right]. \end{aligned} \tag{6.6.6}$$

The backward equation follows from differentiating with respect to t both sides of the Chapman–Kolmogorov equation

$$P(n_1,s|n_0,0)=\sum_n P(n_1,s|n,t)P(n,t|n_0,0).$$

Summing Eq. (6.6.6) from $n=0$ to $n=N-1$ shows that

$$\frac{dS(n_0,t)}{dt}=\omega_+\left[S(n_0+1,t)-S(n_0,t)\right]+\omega_-\left[S(n_0-1,t)-S(n_0,t)\right], \tag{6.6.7}$$

supplemented by the boundary conditions $S(N,t)=0$ and $S(0,t)=S(-1,t)$ and the initial condition $S(n_0,0)=1$.

Let us now calculate the mean elongation time. Integrating Eq. (6.6.7) with respect to t gives

$$-1=\omega_+[T(n_0+1)-T(n_0)]+\omega_-[T(n_0-1)-T(n_0)] \tag{6.6.8}$$

with $T(N)=0$ and $T(0)=T(-1)$. Setting $U(n_0)=T(n_0)-T(n_0-1)$ we obtain the first-order difference equation

$$\omega_+U(n_0+1)-\omega_-U(n_0)=-1,$$

which has the solution $U(0)=0$, $U(1)=-1/\omega_+$, $U(2)=-1/\omega_+-\omega_-/\omega_+^2$, etc. that is

$$U(n)=-\frac{1}{\omega_+}\left[1+\frac{\omega_-}{\omega^+}+\left(\frac{\omega_-}{\omega^+}\right)^2+\ldots+\left(\frac{\omega_-}{\omega^+}\right)^{n-1}\right].$$

Since $-T(n)=U(N)+U(N-1)+\ldots+U(n+1)$, we deduce that

$$T(n_0)=\sum_{n_0+1}^{N}\frac{1}{\omega_+}\sum_{m=0}^{n-1}\left(\frac{\omega_-}{\omega^+}\right)^m. \tag{6.6.9}$$

Introducing $K=\omega_-/\omega_+$ and noting that $0\leq K<1$, we can sum the geometric series to give [204, 661]

$$T(n_0)=\frac{1}{\omega_+}\sum_{n=n_0+1}^{N}\frac{1-K^n}{1-K}=\frac{1}{\omega_+(1-K)}\left[N-n_0-\frac{K^{n_0+1}-K^{N+1}}{1-K}\right].$$

Finally, setting $n_0=0$ we obtain the mean elongation time $\mu=T(0)$ with

$$\mu=\frac{1}{\omega_+(1-K)}\left[N-\frac{K(1-K^N)}{1-K}\right]. \tag{6.6.10}$$

The variance can be calculated in a similar fashion (see Ex. 6.9). Here we simply note that when chain lengthening is dominant, $K\ll 1$, both the mean and variance are linear functions of the chain length N:

$$\mu=\frac{N}{\omega_+}+K\frac{N-1}{\omega_+}+O(K^2), \tag{6.6.11}$$

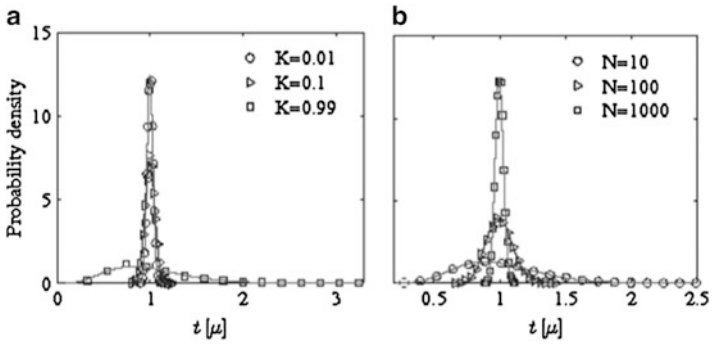


Fig. 6.18: Distribution of elongation times (in units of the mean elongation time) for the TEC model without backtracking. Results from mean-field theory are given by *solid curves* and superimposed with stochastic simulations results. (**a**) Results for $N = 1,000\,\text{bp}$, $\omega_+ = 20\,\text{s}^{-1}$, and various values of $K = \omega_-/\omega_+$. (**b**) Results for $K = 0.01$, $\omega_+ = 20\,\text{s}^{-1}$, and different template lengths N (Adapted from Voliotis et al. [661])

and

$$\sigma^2 = \frac{N}{\omega_+^2} + 4K\frac{N-1}{\omega_+^2} + O(K^2). \tag{6.6.12}$$

For sufficiently long sequences $N \gg 1$, one finds that the distribution of elongation times is given by a narrow Gaussian with fluctuations scaling as $1/\sqrt{N}$. This adds a characteristic delay to the Poisson-like distribution of initiation times (see Fig. 6.18).

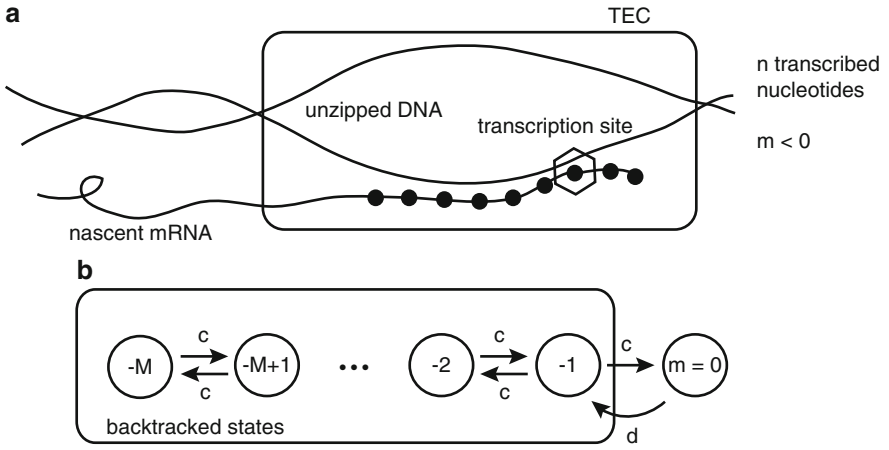


Fig. 6.19: Schematic illustration of TEC backtracking. (**a**) Example of a backtracked state of the TEC with $m = -2$. (**b**) Unbiased random walk model of backtracking. Transitions between backtracked states occur at a rate c. The TEC enters the backtracking regime from the state $m = 0$ at a rate d and exits the backtracking regime at the rate c. (Redrawn from [661].)

Moreover, if initiation is much faster than elongation, then the transcription time is much more regular than if initiation dominates.

Voliotis et al. [661] show that the above picture persists when backtracking pauses are included, provided that they are sufficiently rare. However, the distribution of elongation times is drastically altered when backtracking becomes significant. The simplest way to incorporate backtracking into the model is to treat it as a separate process. That is, one can introduce additional translocation states of the TEC given by $m = -1, \ldots, -M$, which represent backtracked states shifted by $|m|$ steps from state $m = 0$. The duration of backtracking pauses can also be analyzed in terms of a FPT problem—in this case, a random walk on a finite lattice with a reflecting boundary at $m = -M$ and an absorbing boundary at $m = 0$ (see Fig. 6.19). One finds that there is a broad distribution of pause durations that exhibits power law behavior at intermediate duration times. Consequently, the distribution of elongation times is significantly altered. Numerical simulations of the full model also suggest that the distribution of elongation times with long pauses naturally exhibits switching between high and low mRNA product rats, resulting in transcriptional bursting.

6.7 Kinetic Proofreading

A major requirement for proper cell function is that the genetic code is "read" with few mistakes during protein synthesis or DNA replication. For example, both transcription and translation involve the incorporation of specific molecular substrates at particular times, namely, a specific mRNA nucleotide during the production of mRNA or a specific amino acid during production of a protein. The incorporation of each substrate involves some recognition site within an RNAP or a ribosome, respectively, that is more energetically disposed to bind the correct substrate C, say, rather than an incorrect substrate D. In a simple reaction scheme, the frequency of errors is of the order $e^{-\Delta G_{CD}/k_B T}$, where ΔG_{CD} is the smallest difference in binding energies between the correct substrate and an incorrect substrate. The basic problem is that typical values of ΔG_{CD} cannot account for the small error rates observed in protein synthesis. For example, the maximum frequency at which a wrong but similar amino acid is inserted during protein translation is 10^{-4}, which means that even smaller error rates must occur in each recognition step. The error rates are smaller still in the case of DNA transcription, taking values around 10^{-9}. *Kinetic proofreading* is a mechanism for error correction in biochemical processes, which was first introduced by Hopfield [273] and independently by Ninio [479]. The proofreading mechanism increases specificity of biochemical interactions by including a number of intermediate steps that can undo errors at the cost of increased reaction time and free energy expenditure.

Box 6B. Enzyme kinetics.

Enzymes are generally protein catalysts that help convert other molecules called substrates into products, without themselves being changed by the reaction. In contrast to single-step reactions, the rate of reaction does not increase linearly with the concentration of substrate, since it saturates at high concentrations. A simple model to explain this behavior was first proposed by Michaelis and Menten. The basic reaction scheme involves an enzyme E converting a substrate S to a complex C, which then breaks down to form the product P together with the original enzyme. This can be represented by the following two-step process:

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_2} P + E. \tag{6.7.1}$$

Although all the reactions are reversible, reaction rates are typically measured under conditions in which the product P is continually removed from the system, which effectively prevents the final reverse reaction from occurring. Setting $s = [S], c = [C], e = [E]$ and $p = [P]$, we have the system of kinetic equations

$$\frac{ds}{dt} = k_{-1}c - k_1 se, \tag{6.7.2a}$$

$$\frac{de}{dt} = (k_{-1} + k_2)c - k_1 se, \tag{6.7.2b}$$

$$\frac{dc}{dt} = -(k_{-1} + k_2)c + k_1 se, \tag{6.7.2c}$$

$$\frac{dp}{dt} = k_2 c. \tag{6.7.2d}$$

Note that the total concentration of enzyme is conserved, $e + c = e_0$ for some constant e_0. Hence, we can eliminate e such that

$$\frac{dc}{dt} = k_1 e_0 s - (k_{-1} + k_2 + k_1 s)c.$$

If the total concentration of enzyme is small, then s changes relatively slowly, which suggests that c reaches steady state before s changes significantly. Thus,

$$c = e_0 \frac{k_1 s}{k_{-1} + k_2 + k_1 s}.$$

Under this so-called equilibrium approximation the overall rate of generating product (or depleting substrate) is

$$\frac{dp}{dt} = k_2 e_0 \frac{k_1 s}{k_{-1} + k_2 + k_1 s} = k_2 e_0 \frac{s}{s + K_M}, \tag{6.7.3}$$

where K_M is the Michaelis constant

$$K_M = \frac{k_{-1} + k_2}{k_1}.$$

Hence, the rate of production is linear s when the substrate concentration is low but saturates when s is sufficiently large. The resulting behavior is referred to as Michaelis–Menten kinetics. For a more general discussion of various kinetic schemes including Michaelis–Menten, see the books by Siegel and Edelstein-Keshet [579] and Keeener and Sneyd [322].

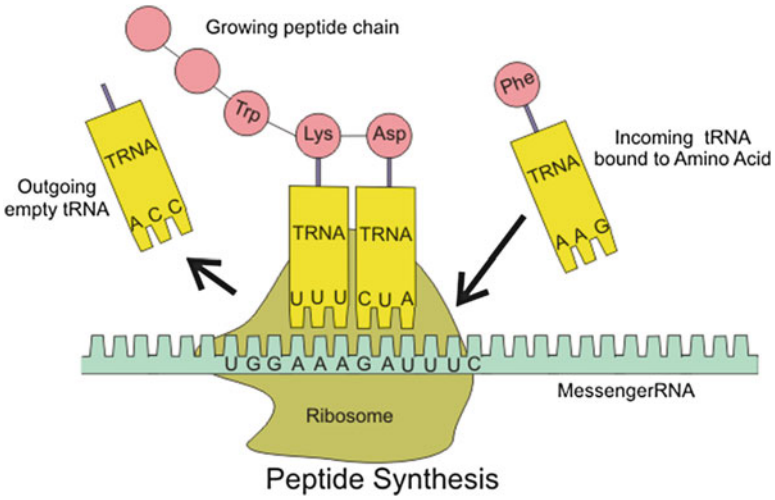


Fig. 6.20: Ribosomes can bind to an mRNA chain and use it as a template for determining the correct sequence of amino acids in a particular protein. Amino acids are selected, collected, and carried to the ribosome by transfer RNA (*tRNA*) molecules, which enter one part of the ribosome and bind to the messenger RNA chain. The attached amino acids are then linked together by another part of the ribosome. Once the protein is produced, it can then "fold" to produce a specific functional three-dimensional structure. Specificity is achieved through the interaction between the codon (triplet of nucleotides) in mRNA and the anti-codon in the tRNA. (Public domain figure from Wikipedia)

6.7.1 Kinetic Proofreading in Protein Synthesis

Consider the binding interaction between a codon E of mRNA and the anti-codon of a tRNA during protein synthesis (see Fig. 6.20). Let C denote the correct tRNA and D an incorrect tRNA. Here E may be viewed as an enzyme acting on a substrate C or D according to a classical *Michaelis–Menten* scheme (see Box 6B). That is,

$$E \underset{k_{\text{off}}}{\overset{k_{\text{on}}[C]}{\rightleftharpoons}} EC \xrightarrow{W} E + \text{correct amino acid incorporated} \tag{6.7.4a}$$

$$E + D \underset{k'_{\text{off}}}{\overset{k'_{\text{on}}[D]}{\rightleftharpoons}} ED \xrightarrow{W} E + \text{incorrect amino acid incorporated.} \tag{6.7.4b}$$

(It is assumed that the catalytic step has no selectivity, that is, the rate of catalysis W is the same for both substrates.) The corresponding kinetic equations are

$$\frac{d[EC]}{dt} = k_{\text{on}}[E][C] - (k_{\text{off}} + W)[EC] \tag{6.7.5a}$$

$$\frac{d[ED]}{dt} = k'_{\text{on}}[E][D] - (k'_{\text{off}} + W)[ED], \tag{6.7.5b}$$

$$[E]_{\text{Total}} = [EC] + [ED] + [E]. \tag{6.7.5c}$$

The last equation ensures that the total concentration of ribosomes or enzymes is fixed. At steady state, we have

$$[EC] = [E]\frac{k_{\text{on}}[C]}{k_{\text{off}} + W}, \quad [ED] = [E]\frac{k'_{\text{on}}[D]}{k'_{\text{off}} + W}.$$

It follows that the rates of correct and incorrect translation are

$$R_{\text{correct}} = W[EC], \quad R_{\text{incorrect}} = W[ED],$$

and the error rate is

$$F_0 = \frac{R_{\text{incorrect}}}{R_{\text{correct}}} = \left[\frac{k'_{\text{on}}[D]}{k'_{\text{off}} + W}\right]\left[\frac{k_{\text{on}}[C]}{k_{\text{off}} + W}\right]^{-1}. \tag{6.7.6}$$

Typically, one finds that the on rates are approximately the same for all tRNAs (being diffusion limited) and the tRNAs have similar concentrations, that is, $k_{\text{on}} \approx k'_{\text{on}}$ and $[D] \approx [C]$. Hence, the error rate F_0 is minimized by taking the catalytic rate W to be much smaller than the off rates. Introducing the dissociation constants $K_C = k_{\text{off}}/k_{\text{on}}, K_D = k'_{\text{off}}/k'_{\text{on}}$, we have

$$F_0 \approx \frac{K_C}{K_D} = e^{-\Delta G_{CD}/k_B T}. \tag{6.7.7}$$

The above simple binding model neglects the fact that when tRNA binds to a codon, it is chemically altered via the hydrolysis of GTP; an analogous process occurs during the polymerization of microtubules (see Sect. 4.1). The transition to the new state is irreversible and in this state the tRNA can also dissociate from the mRNA. This leads to the new reaction scheme

$$E \underset{k_{\text{off}}}{\overset{k_{\text{on}}[C]}{\rightleftharpoons}} EC \xrightarrow{r} EC^* \xrightarrow{W} E + \text{correct amino acid incorporated} \qquad (6.7.8a)$$
$$\downarrow E$$

$$E \underset{k'_{\text{off}}}{\overset{k'_{\text{on}}[D]}{\rightleftharpoons}} ED \xrightarrow{r} ED^* \xrightarrow{W} E + \text{incorrect amino acid incorporated.} \qquad (6.7.8b)$$
$$\downarrow E$$

Let the on and off rates of the modified substrates C^* and D^* be $q_{\text{off}}, q_{\text{on}}[C^*] \approx 0$ and $q'_{\text{off}}, q'_{\text{on}}[D^*]$, respectively, with $[C^*], [D^*] \approx 0$. The steady-state concentrations of the modified substrate then satisfy (see Ex. 6.10),

$$[EC^*] = \frac{1}{q_{\text{off}} + W} \frac{r k_{\text{on}}}{k_{\text{off}} + r} [E][C].$$

Again assuming that the rates of catalysis W, r are much smaller than the on and off rates, and taking the concentrations of all tRNAs to be the same, we obtain the new error rate

$$F \approx \frac{q_{\text{off}}}{q'_{\text{off}}} \frac{k'_{\text{on}}}{k'_{\text{off}}} \frac{k_{\text{off}}}{k_{\text{on}}}.$$

Finally, taking the on rates to be tRNA nonspecific,

$$F = \frac{Q_C}{Q_D} \frac{K_C}{K_D} = e^{-\Delta G_{CD}/k_B T} e^{-\Delta G_{C^*D^*}/k_B T} < F_0. \qquad (6.7.9)$$

In particular, if the difference in biding energies of the two substrates is the same for the native and modified states, then

$$F = \left[e^{-\Delta G_{CD}/k_B T} \right]^2. \qquad (6.7.10)$$

In summary, the inclusion of an irreversible step into the kinetic scheme, $EA \to EA^*$, which necessitates the expenditure of energy, provides an additional opportunity for the incorrect substrate to dissociate and leads to a reduction in the error rate. An even higher level of accuracy can be achieved by having a sequence of n irreversible proofreading stages:

$$E \underset{k_{\text{off}}}{\overset{k_{\text{on}}[C]}{\rightleftharpoons}} EC \xrightarrow{r_1} EC^* \xrightarrow{r_2} EC^{**} \cdots \xrightarrow{r_m} EC^{**\cdots *} \xrightarrow{W} E + \text{product}.$$
$$\downarrow E \qquad \downarrow E \qquad \downarrow E$$

6.7.2 Kinetic Proofreading in T-Cell Activation

T cells, which mature in the thymus, are one of two key cell types of the adaptive immune system, whose basic function is the detection and destruction of intracellular pathogens such as certain bacteria and all viruses (see the review by Coombs and Goldstein [125]). (The other cell type consists of B cells, which mature in the bone marrow and are mainly concerned with the detection and destruction of extracellular pathogens.) In order to execute their function, T cells scan the surfaces of cells for molecular markers of infection. Detection of the appropriate marker activates the T cell which then responds to the pathogen, either by killing the infected cell (effector T cells) or by signaling other parts of the immune system such as B cells (helper T cells). Since T cells only scan the surface of other cells, it is necessary that some cells are able to present information regarding their internal contents to the surface. This is achieved by cutting intracellular proteins into peptide fragments and transporting these fragments to the surface for surveillance by T cells. If a pathogen is present within the cell, then signature peptide groups known as antigens will be made accessible. A major challenge for the pathogen recognition machinery is that the vast majority of peptides on a given antigen-presenting cell do not signify the presence of a pathogen. Thus, a T cell has to recognize an antigen against a noisy background of these so-called self-peptides, just as a ribosome has to recognize the correct tRNA during each stage of protein synthesis. It is not too surprising, therefore, that a kinetic proofreading model has been developed for T-cell activation by McKeithan [430].

The model of McKeithan considers the interaction of a T-cell receptor (TCR) with a ligand consisting of a peptide fragment that is bound to a specialized molecule in the surface of an antigen-presenting cell, known as a major histocompatibility complex (MHC) molecule (see Fig. 6.21a). The peptide–MHC complex that forms the ligand is denoted by pMHC. There are two basic assumptions of the model: (i) In order to respond to an antigen, a TCR in an inactive state T has to undergo a sequence of N modifications to form the activated state B_N. (ii) Dissociation of pMHC from the TCR can occur at any stage, after which the receptor quickly returns to its inactive state (see Fig. 6.21b). Suppose that the off rate back to the inactive state T is the same for all intermediate states. We then have the following hierarchy of kinetic equations for the concentrations $[T], [B_j]$, $j = 0, \ldots, N$:

$$\frac{d[T]}{dt} = -k_{\text{on}}[T][P] + k_{\text{off}} \sum_{i=0}^{N} [B_i], \tag{6.7.11a}$$

$$\frac{d[B_0]}{dt} = k_{\text{on}}[T][P] - k_{\text{off}}[B_0] - k_p[B_0], \tag{6.7.11b}$$

$$\frac{d[B_i]}{dt} = k_p([B_{i-1}] - [B_i]) - k_{\text{off}}[B_i], \tag{6.7.11c}$$

$$\frac{d[B_N]}{dt} = k_p[B_{N-1}] - k_{\text{off}}[B_N], \tag{6.7.11d}$$

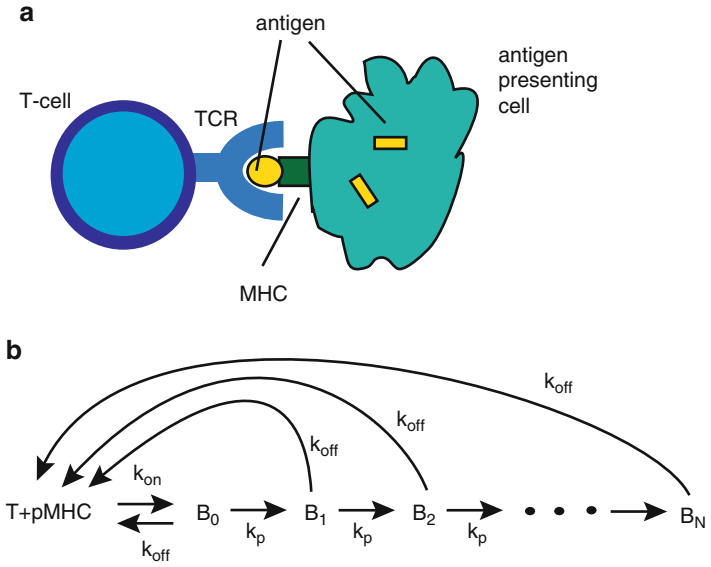


Fig. 6.21: Kinetic proofreading model of T-cell activation. (**a**) Schematic diagram of a T-cell receptor (*TCR*) binding to an antigen that is attached to a major histocompatibility complex molecule (*MHC*) in the surface of an antigen-presenting cell. (**b**) Reaction diagram, see text for details

where $[P]$ is the concentration of a specific pMHC complex. Solving these equations in steady state shows that the fraction of activated complexes is (see Ex. 6.10)

$$\frac{[B_N]}{\sum_{i=0}^{N}[B_i]} = \left(\frac{k_p}{k_p + k_{\text{off}}}\right)^N . \tag{6.7.12}$$

Note that $k_p/(k_p + k_{\text{off}})$ is the probability that in any intermediate step i, the T cell is modified before dissociation of the pMHC. Assuming that k_p is independent of the particular substrate, it follows that the off rate k_{off} is the only parameter whose variation can distinguish between peptides. Even for small values of N, the fraction of activated T cells is sensitive to small changes in k_{off}, reflecting the objective of the kinetic proofreading mechanism. However, it comes at a cost, namely that the actual value of the activity level in response to the correct antigen reduces as N increases, so that an increase in selectivity coincides in a decrease in sensitivity.

6.8 Stochastic Algorithms for Chemical Kinetics

6.8.1 The Stochastic Simulation Algorithm

The SSA, which was originally developed by Gillespie [217–219], is an efficient numerical scheme for generating exact sample paths of a continuous-time Markov process whose probability distribution evolves according to a chemical master equation. Following Sect. 6.2, suppose that the mass-action kinetics of a general biochemical network is written in the form

$$\frac{dx_i}{dt} = \sum_{a=1}^{R} S_{ia} f_a(\mathbf{x}), \quad i = 1,\ldots,N, \tag{6.8.1}$$

where a labels a single-step reaction, f_a are the transition intensities or propensities, and $\mathbf{S}$ is the $N \times R$ stochiometric matrix. Given this notation, the corresponding master equation is

$$\frac{dP(\mathbf{n},t)}{dt} = \Omega \sum_{a=1}^{R} \left(\prod_{i=1}^{N} \mathbb{E}^{-S_{ia}} - 1 \right) f_a(\mathbf{n}/\Omega) P(\mathbf{n},t), \tag{6.8.2}$$

where Ω represents the system size. Typically, Ω is the volume of the well-mixed compartment where reactions occur or the total number of molecules in cases where there is number conservation. Here $\mathbb{E}^{-S_{ia}}$ is a step or ladder operator such that for and function $g(\mathbf{n})$,

$$\mathbb{E}^{-S_{ia}} g(n_1,\ldots,n_i,\ldots,n_N) = g(n_1,\ldots,n_i - S_{ia},\ldots,n_N). \tag{6.8.3}$$

In the following we eliminate the global factor of Ω by rescaling time $t \to \Omega t$.

The starting point for constructing the SSA is to define a new probability function $p(\tau,a|\mathbf{x},t)$, which is the probability, given $\mathbf{X}(t) = \mathbf{x}$, that the next reaction in the system will occur in the time interval $[t+\tau, t+\tau+\Delta\tau)$ and will be the reaction a. From this perspective, both τ and a are random variables conditioned on $\mathbf{X}(t) = \mathbf{x}$. An analytical expression for $p(\tau,a|\mathbf{x},t)$ can be obtained by introducing another probability function $P_0(\tau|\mathbf{x},t)$, which is the probability, given $\mathbf{X}(t) = \mathbf{x}$, that no reaction of any kind occurs in the time interval $[t,t+\tau)$. It follows from the definitions of P_0 and the propensities f_a that P_0 satisfy the equation

$$P_0(\tau + d\tau|\mathbf{x},t) = P_0(\tau|\mathbf{x},t) \left[1 - \sum_{a=1}^{R} f_a(\mathbf{x}) d\tau \right],$$

which is the product of the probability that no reaction occurs in $[t,\tau)$ and the probability that there are no transitions in the infinitesimal interval $[t+\tau, t+\tau+d\tau)$. Rearranging and taking the limit $d\tau \to 0$ yields

$$\frac{dP_0(\tau|\mathbf{x},t)}{d\tau} = -F(\mathbf{x}) P_0(\tau|\mathbf{x},t), \quad F(\mathbf{x}) = \sum_{a=1}^{R} f_a(\mathbf{x}).$$

Under the initial condition $P(0|x,t) = 1$, we have the solution

$$P_0(\tau|\mathbf{x},t) = \exp(-F(\mathbf{x})\tau).$$

We now note

$$p(\tau,a|\mathbf{x},t)d\tau = P_0(\tau|\mathbf{x},t)f_a(\mathbf{x})d\tau,$$

which implies that p can be written in the form

$$p(\tau,a|\mathbf{x},t) = F(\mathbf{x})\exp(-F(\mathbf{x})\tau)\frac{f_a(\mathbf{x})}{F(\mathbf{x})}. \tag{6.8.4}$$

Hence, τ is an exponential random variable with mean and standard deviation $1/F(\mathbf{x})$, while a is a statistically independent integer random variable with $\mathbf{x}$-dependent probability $f_a(\mathbf{x})/F(\mathbf{x})$.

One exact Monte Carlo method for generating samples of the random variables τ, a is to draw two random numbers r_1, r_2 from the uniform distribution on $[0,1]$ and take

$$\tau = -\frac{1}{F(\mathbf{x})}\ln r_1 \tag{6.8.5a}$$

$$a = \text{ the smallest integer for which } \sum_{s=1}^{a} f_a(\mathbf{x}) > r_2 F(\mathbf{x}). \tag{6.8.5b}$$

The direct method of implementing the SSA is as follows:

1. Initialize the time $t = t_0$ and the chemical state $\mathbf{x} = \mathbf{x}_0$.
2. Given the state $\mathbf{x}$ at time t, determine the $f_a(\mathbf{x})$ for $a = 1,\ldots,R$ and their sums $F(\mathbf{x})$.
3. Generate values for τ and a using Eq. (6.8.5).
4. Implement the next reaction by setting $t \to t' = t + \tau$ and $x_j \to x_j' = x_j + S_{ja}/\Omega$.
5. Return to step 2 with $(\mathbf{x},t)$ replaced by $(\mathbf{x}',t')$, or else stop.

There have been a variety of subsequent algorithms that differ in the implementation of step 2, including the next reaction method [215] and the modified next reaction method [8]. The latter is based on the random time-change representation of Kurtz, which will be considered in Chap. 11 after developing the theory of martingales.

6.8.2 Tau-Leaping

In many applications the mean time between reactions, $1/F(\mathbf{x})$, is very small so that simulating every reaction becomes computationally infeasible, irrespective of the version of the SSA chosen. Gillespie [218] introduced *tau-leaping* in order to

address this problem by sacrificing some degree of exactness of the SSA in return for a gain in computational efficiency. The basic idea is to "leap" the system forward by a pre-selected time τ (distinct from the τ of the SSA), which may include several reaction events. Given $\mathbf{X}(t) = \mathbf{x}$, τ is chosen to be large enough for efficient computation but small enough so that

$$f_a(\mathbf{x}) \approx \text{ constant in } [t, t+\tau) \text{ for all } a.$$

Let $\mathscr{N}(\lambda)$ denote a Poison counting process with mean λ. During the interval $[t, t+\tau)$ there will be approximately $\mathscr{N}(\lambda_a)$ reactions of type a with $\lambda_a = f_a(\mathbf{x})\tau$. Since each of these reactions increases x_j by S_{ja}/Ω, the state at time $t+\tau$ will be

$$X_j(t+\tau) = \mathbf{x} + \sum_{a=1}^{R} \mathscr{N}_a(f_a(\mathbf{x})\tau) S_{ja}, \tag{6.8.6}$$

where the $\mathscr{N}_a$ are independent Poisson processes. This equation is known as the *tau-leaping formula*. However, there are two fundamental problems with the original formulation of tau-leaping. First, it is difficult to choose the appropriate value of τ at each iteration of the algorithm—occasionally large changes in propensities occur that cause one or more components x_j to become negative. Second, although tau-leaping becomes exact in the limit $\tau \to 0$, the inefficiency becomes prohibitive since the R generated Poisson random numbers will be zero most of the time resulting in no change of state. These two issues have been addressed in various modifications in the tau-leaping procedure (see for example Cao et al. [93]).

6.9 Exercises

Problem 6.1 (Bursting in protein translation I).

(a) Consider a single mRNA molecule which produces n proteins with probability

$$P(n) = \left(\frac{r}{r+\gamma}\right)^n \frac{\gamma}{r+\gamma}.$$

Use a generating function to show that the burst size $b \equiv \langle n \rangle = r/\gamma$.

(b) Calculate the Laplace transform

$$\tilde{P}(s) = \int_0^\infty P(n) \mathrm{e}^{-ns} dn$$

with n treated as a continuous variable (for large protein number).

(c) Evaluate the inverse Laplace transform of $\tilde{P}_m(s) = [\tilde{P}(s)]^m$ to obtain the result

$$P_m(n) = \left(\frac{b}{1+b}\right)^n \left(\frac{1}{1+b}\right)^m \frac{n^{m-1}}{\Gamma(m)}.$$

Problem 6.2 (Bursting in protein translation II). Consider the Chapman–Kolmogorov equation (6.2.8) for protein bursting:

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial}{\partial x}[\gamma_0 x p(x)] + k\int_0^x w(x-x')c(x')p(x',t)dx',$$

with

$$w(x) = \frac{1}{b}e^{-x/b} - \delta(x).$$

(a) Suppose that $c(x) = 1$ (no autoregulatory feedback). Laplace transforming the steady-state equation with respect to the protein number x, show that the stationary distribution is given by the gamma distribution

$$p(x) = \frac{1}{b^m \Gamma(m)} x^{m-1} e^{-x/b}, \quad m = \frac{k}{\gamma_0}.$$

(For a more challenging problem, Laplace transform the full time-dependent equation, solve the resulting quasilinear PDE in Laplace space using the method of characteristics, and show that the system converges to the gamma distribution in the limit $t \to \infty$.)

(b) Suppose that $c(x)$ is given by the Hill function

$$c(x) = \frac{k^s}{k^s + x^s} + \varepsilon.$$

Using Laplace transforms along similar lines to part (a), show that the stationary probability density is

$$p(x) = Ax^{m(1+\varepsilon)-1} e^{-x/b}[1 + (x/k)^s]^{-m/s},$$

where A is a normalization factor.

(c) Plot the stationary density of part (b) for the parameter values $m = 10, b = 20,$ and $k = 70nM$ and the following four cases: (i) $c \equiv 1$ (no feedback); (ii) $s = +1$, $\varepsilon = 0.05$; (iii) $s = -1, \varepsilon = 0.2$; (iv) $s = -4, \varepsilon = 0.2$. Hence show that negative feedback reduces noise, whereas positive feedback enhances noise and can lead to bistability.

Problem 6.3 (Binary response in stochastic gene expression). Consider the stochastic model of a gene expression in which the gene randomly switches between an active and inactive state. The steady-state probability densities $p_{0,1}(x)$ for protein concentration x when the gene is in an active ($j = 1$) or inactive ($j = 0$) state satisfy the pair of equations

$$\frac{d}{dx}(-\gamma x p_0(x)) = k_- p_1(x) - k_+ p_0(x)$$
$$\frac{d}{dx}([r - \gamma x]p_1(x)) = k_+ p_0(x) - k_- p_1(x)$$

with boundary conditions $p_0(r/\gamma) = 0$ and $p_1(0) = 0$.

(a) Derive the normalization conditions

$$\int_0^{r/\gamma} p_0(x)dx = \frac{k_-}{k_- + k_+}, \quad \int_0^{r/\gamma} p_1(x)dx = \frac{k_+}{k_- + k_+}.$$

(b) By adding the pair of steady-state equations show that one solution is

$$p_0(x) = \frac{r - \gamma x}{\gamma x} p_1(x).$$

(c) Substituting for $p_0(x)$, solve the resulting differential equation for $P(x) = (r - \gamma x)p_1(x)$, and thus obtain the solution

$$p_0(x) = C(\gamma x)^{-1+k_+/\gamma}(r - \gamma x)^{k_-/\gamma}, \quad p_1(x) = C(\gamma x)^{k_+/\gamma}(r - \gamma x)^{-1+k_-/\gamma}.$$

(d) Using part (c), show that

$$\int_0^{r/\gamma} p_0(x)dx = \frac{C}{\gamma} r^{(k_+ + k_-)/\gamma} B(k_+/\gamma, 1 + k_-/\gamma),$$
$$\int_0^{r/\gamma} p_1(x)dx = \frac{C}{\gamma} r^{(k_+ + k_-)/\gamma} B(1 + k_+/\gamma, k_-/\gamma),$$

where $B(\alpha, \beta)$ is the beta function:

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt.$$

(e) Using the standard property

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)},$$

show that the solution in part (c) satisfies the normalization conditions provided that

$$C = \gamma \left[r^{(k_+ + k_-)/\gamma} B(k_+/\gamma, k_-/\gamma) \right]^{-1}.$$

Problem 6.4 (Linear noise approximation of a two-state gene regulatory network). Consider the simple kinetic model of gene expression given by equations

$$\frac{dx_1}{dt} = k_+(1 - x_1) - k_- x_1, \quad \frac{dx_2}{dt} = r x_1 - \gamma x_2.$$

Here x_1 is the density of active genes and x_2 is the density of protein. There is a unique fixed point

$$x_1^* = \frac{k_+}{k_+ + k_-} x_{\max}, \quad x_2^* = \frac{r}{\gamma} x_1^*.$$

Applying the linear noise approximation to the corresponding master equation for finite copy numbers yields an OU process whose stationary covariance matrix Σ satisfies the matrix equation

$$\mathbf{A}\Sigma + \Sigma\mathbf{A}^T = -\mathbf{B}\mathbf{B}^T \equiv -\mathbf{D},$$

with

$$\mathbf{A} = \begin{pmatrix} -(k_+ + k_-) & 0 \\ r & -\gamma \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} 2k_- x_1^* & 0 \\ 0 & 2rx_1^* \end{pmatrix}.$$

By solving the matrix equation in component form, determine the variances σ_1^2 and σ_2^2 for $Y_i = (X_i - x_i^*)/\sqrt{\Omega}$, where Ω is the system size.

Problem 6.5 (Frequency domain analysis of a simple gene network). Consider a simple model of protein translation given by the stochastic kinetic equations

$$\frac{dx}{dt} = k - \gamma x + \eta(t), \quad \frac{dy}{dt} = rx - \gamma_p y + \eta_p(t),$$

where x and y are concentrations of mRNA and protein, γ, γ_p are degradation rates, k is the rate of mRNA production, and r is the rate of protein production. Moreover $\eta(t)$ and $\eta_p(t)$ are independent white noise terms with $\langle \eta \rangle = \langle \eta_p \rangle = 0$, and

$$\langle \eta(t)\eta(t') \rangle = q\delta(t-t'), \quad \langle \eta_p(t)\eta_p(t') \rangle = q_p\delta(t-t'), \quad \langle \eta(t)\eta_p(t') \rangle = 0.$$

(a) Introducing the Fourier transforms

$$\tilde{\eta}(\omega) = \int_{-\infty}^{\infty} \mathrm{e}^{i\omega t}\eta(t)dt, \qquad \eta(t) = \int_{-\infty}^{\infty} \mathrm{e}^{-i\omega t}\tilde{\eta}(\omega)\frac{d\omega}{2\pi},$$

show that

$$\langle \eta(\omega)\eta(\omega') \rangle = 2q\pi\delta(\omega + \omega').$$

(b) By linearizing about the steady state $x^* = k/\gamma$, $y^* = rk/(\gamma\gamma_p)$ and using Fourier transforms show that the power spectra of the fluctuations $X(t) = x(t) - x^*$ and $Y(t) = y(t) - y^*$ are given by

$$S_{XX}(\omega) = \frac{q}{\omega^2 + \gamma^2}, \quad S_{YY}(\omega) = \frac{q_p}{\omega^2 + \gamma_p^2} + \frac{r^2 q}{(\omega^2 + \gamma^2)(\omega^2 + \gamma_p^2)}.$$

(c) Using the definition of the power spectrum, written in the form

$$\langle X(t)^2 \rangle = \int_{-\infty}^{\infty} S_{XX}(\omega)\frac{d\omega}{2\pi},$$

show that

$$\langle X(t)^2 \rangle = \frac{q}{2\gamma}.$$

Similarly, show that

$$\langle Y(t)^2\rangle = \frac{q_p}{2\gamma_p} + \frac{r^2 q}{2\gamma_p \gamma^2} + \mathcal{O}(\gamma^{-3}).$$

Hint: you should assume that $\gamma \gg \gamma_p$ and use the result

$$\int_{-\infty}^{\infty} \frac{1}{\omega^2 + a^2} \frac{d\omega}{2\pi} = \frac{1}{2a}.$$

(d) From the linear noise approximation, one obtains the following Fano factors for mRNA (m) and proteins (n):

$$\frac{\mathrm{var}[m]}{\langle m\rangle} = 1, \quad \frac{\mathrm{var}[n]}{\langle n\rangle} = 1 + b,$$

where $b = r/\gamma$. Use this to determine q and q_p.

Problem 6.6 (Attenuation of noise in signaling cascades). Consider a generic model of a stochastic signaling cascade consisting of molecular species labeled $i = 0, \ldots, n$ with corresponding concentrations y_i [624] (see Fig. 6.22). Suppose that the rate of production of species i only depends on the concentration y_{i-1} of the species at the previous level of the cascade and that it degrades at a fixed rate γ_i. In the deterministic limit, we have a system of first-order kinetic equations:

$$\dot{y}_i + \gamma_i y_i = f_{i-1}(y_{i-1}),$$

where f_{i-1} is the corresponding production rate function and $f_{-1} = 0$. Suppose that there exists a unique stable steady state. Linearizing about the steady state and adding white noise terms to take into account intrinsic fluctuations, we have the system of linear equations

$$\delta\dot{y}_i + \gamma_i \delta y_i = c_{i-1}\delta y_{i-1} + \eta_i,$$

where

$$\langle \eta_i\rangle = 0, \quad \langle \eta_i(t)\eta_j(t')\rangle = q_i \delta_{i,j}\delta(t - t'),$$

and c_i is the derivative of f_i at the steady state (with $c_{-1} = 0$). That is, c_i can be interpreted as a differential gain or amplification factor. Both γ_i and c_i have units of t^{-1}. For convenience, set $\gamma_i = 1$ for all i.

(a) Using Fourier transforms show that

$$\langle \delta y_n^2(\omega)\rangle = \alpha_n + \beta_{n-1}\alpha_{n-1} + \beta_{n-1}\beta_{n-2}\alpha_{n-2} + \ldots + \beta_{n-1}\ldots\beta_0\langle \delta y_0^2(\omega)\rangle,$$

where

$$\alpha_j = \frac{q_j}{1+\omega^2}, \quad \beta_j = \frac{c_j^2}{1+\omega^2}.$$

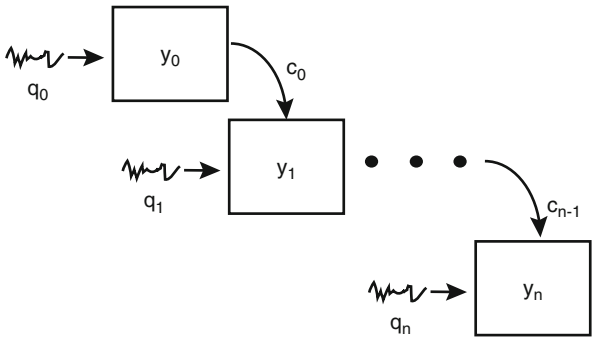


Fig. 6.22: Schematic diagram of a linearized stochastic cascade

(b) Let $q = \max_i\{q_i\}$, $c = \max_i\{c_i\}$ and define

$$\alpha = \frac{q}{1+\omega^2}, \quad \beta = \frac{c^2}{1+\omega^2}.$$

Show that

$$\lim_{n\to\infty}\langle \delta y_n^2(\omega)\rangle \le \frac{\alpha}{1-\beta}.$$

Taking the inverse Fourier transform, obtain the result

$$\lim_{n\to\infty}\langle \delta y_n^2(t)\rangle \le q\int_{-\infty}^{\infty}\frac{1}{1+\omega^2-c^2}\frac{d\omega}{2\pi} = \frac{q}{2\sqrt{1-c^2}}.$$

This establishes that fluctuations in the output of the signaling cascaded will be bounded provided that $|c_i| \le |c| < 1$.

(c) Now consider a finite cascade of length n with $c_i = c < 1$ for all $i = 0,\ldots,n$, $q_i = q$ for all $i = 1,\ldots,n$, and $q_0 > q$. Thus the noise at the input level is higher than in successive levels of the cascade. Using part (a), show that

$$\langle \delta y_n^2(\omega)\rangle = q\sum_{j=0}^{n-1}\frac{c^{2j}}{(1+\omega^2)^{j+1}} + q_0\frac{c^{2n}}{(1+\omega^2)^{n+1}}.$$

Taking inverse Fourier transforms and using contour integration, establish that

$$\langle \delta y_n^2(t)\rangle \le q\sum_{j=0}^{n-1}\frac{(2j)!}{j!\,j!}\frac{c^{2j}}{2^{2j+1}} + q_0\frac{(2n)!}{n!n!}\frac{c^{2n}}{2^{2n+1}}.$$

(d) Applying Stirling's approximation to part (c),

$$j! \approx \left(\frac{j}{e}\right)^j\sqrt{2\pi j},$$

obtain the inequality

$$\langle \delta y_n^2(t) \rangle \leq \frac{q}{2}\left(1+\sum_{j=1}^{n-1}\frac{c^{2j}}{\sqrt{\pi j}}\right)+\frac{q_0}{2}\frac{c^{2n}}{\sqrt{\pi n}}.$$

The first term represents an increase in intrinsic fluctuations with cascade length n, whereas the second represents a faster than exponential decrease in the input noise with cascade length. Show that the optimal cascade length for minimizing the total noise is

$$n_{\text{opt}} = \left[\frac{1}{1-(q_0-q)^2/(q_0^2 c^4)}\right]_-,$$

with $[x]_-$ denoting the greatest integer less than x.

Problem 6.7 (Linear noise approximation of autoregulation). Consider a simple kinetic model of gene autoregulation given by

$$\frac{dx_1}{dt} = -\gamma x_1 + F(x_2), \quad \frac{dx_2}{dt} = r x_1 - \gamma_p x_2,$$

with $F(x) = k_0 - kx$. Here x_1 is the concentration of mRNA and x_2 is the concentration of protein. There is a unique fixed point

$$x_1^* = \frac{k_0 \gamma_p}{\gamma\gamma_p + kr}, \quad x_2^* = \frac{r}{\gamma_p} x_1^*.$$

Applying the linear noise approximation to the corresponding master equation for finite copy numbers yields an OU process whose stationary covariance matrix $\boldsymbol{\Sigma}$ satisfies the matrix equation

$$\mathbf{A}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{A}^T = -\mathbf{B}\mathbf{B}^T \equiv -\mathbf{D},$$

with

$$\mathbf{A} = \begin{pmatrix} -\gamma & -k \\ r & -\gamma_p \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} kx_2^* + \gamma x_1^* & 0 \\ 0 & r x_1^* + \gamma_p x_2^* \end{pmatrix}.$$

Solving the matrix equation show that the Fano factor for proteins is

$$\frac{\text{var}[n]}{\langle n \rangle} = 1 + \frac{b}{1+\eta}\left(1 - \frac{\phi}{1+b\phi}\right),$$

where $b = r/\gamma, \eta = \gamma_p/\gamma, \phi = k/\gamma_p$.

Problem 6.8 (Mutual repressor model). Consider the mutual repressor model with a single promoter site, whose stochastic version is described by the master equation (6.4.6).

(a) Suppose that the kinetics of promoter transitions are much faster than the production and degradation of proteins. Write down the continuous-time Markov equations describing the evolution of the probabilities p_j, assuming the concentrations x, y of proteins X, Y are fixed. (Recall that the proteins bind to the promoter site as dimers.) Solve for the steady-state probabilities p_j^* in terms of x and y.

(b) Under the adiabatic approximation $p_j = p_j^*(x,y)$, with $p_0^* + p_1^*$ ($p_0^* + p_2^*$) interpreted as the rate of production of protein x (y), write down the kinetic equations for x, y and use the solutions of part (a) to derive the deterministic equations

$$\frac{dx}{dt} = f(x,y), \quad \frac{dy}{dt} = f(y,x), \quad \text{with } f(x,y) = \frac{1}{1+\frac{y^2}{b+x^2}} - x.$$

Use phase-plane analysis to construct a bifurcation diagram for this planar system with b treated as a bifurcation parameter.

(c) Derive the Chapman–Kolmogorov equation (6.4.10) by carrying out a system-size expansion of the master equation expressed in the form (6.4.6).

Problem 6.9 (Model of transcriptional elongation). Consider the birth–death master equation for the elongation phase of transcription in the absence of backtracking (Sect. 6.6).

(a) Starting from the backward master equation (6.6.7) derive a difference equation for the second moment $T_2(n_0)$ of the elongation time, where n_0 is the starting position along the chain, analogous to the difference equation (6.6.8) for the first moment.

(b) Solve the difference equation in part (a) recursively by introducing the variable $U_2(n_0) = T_2(n_0) - T_2(n_0 - 1)$.

(c) Using the result from part (b) and the formula (6.6.10) for the mean elongation time, determine the variance σ^2 of the elongation time in terms of $K = \omega_-/\omega_+$ and ω_+, where $\omega_\pm$ are the effective polymerization/depolymerization rates, and show that when $K \ll 1$,

$$\sigma^2 = \frac{N}{\omega_+^2} + 4K\frac{N-1}{\omega_+^2} + O(K^2).$$

Problem 6.10 (Kinetic proofreading). Consider the kinetic proofreading model given by the modified Michaelis–Menten reaction kinetics of Eq. (6.7.8).

(a) Write down the kinetic equations for the evolution of the concentrations $[EC]$ and $[EC^*]$.

(b) Show that the steady-state concentration of the modified enzyme–substrate complex $[EC^*]$ is

$$[EC^*] = \frac{1}{q_{\text{off}} + W}\frac{rk_{\text{on}}}{k_{\text{off}} + r}[E][C].$$

(c) Repeating the analysis for the incorrect substrate D, derive the following approximation for the error rate:

$$F = \frac{Q_C}{Q_D}\frac{K_C}{K_D} = e^{-\Delta G_{CD}/k_BT} e^{-\Delta G_{C^*D^*}/k_BT}.$$

(d) Now consider the kinetic proofreading model of T-cell activation given by Eq. (6.7.11). Show that the steady-state fraction of active TCRs is

$$\frac{[B_N]}{\sum_{i=0}^{N}[B_i]} = \left(\frac{k_p}{k_p + k_{\rm off}}\right)^N.$$

Problem 6.11 (Computer simulations: gene networks). Write MatLab programs based on the Gillespie algorithm (see Sect. 6.8) that generate trajectories for each of the following gene networks.

(a) The model of regulated transcription whose master equation is given by Eq. (6.3.9). There are two discrete variables (number of active genes n_1 and number of mRNA molecules n_2) and four reactions (gene activation and deactivation, mRNA production and degradation). Take the parameter values $k_+ = 0.03\,\text{min}^{-1}, k_- = 0.2\,\text{min}^{-1}, k = 10\,\text{min}^{-1}$, and $\gamma = 0.2\,\text{min}^{-1}$ and consider the two cases $n_{\max} = 10$, and $n_{\max} = 100$. Run the simulations for sufficient time to reach steady state. Plot a histogram of $n_1(T)$ and $n_2(T)$ based on 100 simulations, say, where T is the final time. Determine the mean and variance, and compare the numerical Fano factor with the theoretical expressions based on the diffusion approximation.

(b) The mutual repressor model whose master equation is given by Eq. (6.4.3). You will have to determine the stochiometric matrix and the propensities. There are three discrete variables (number of X proteins n, number of Y proteins m, and state of the promoter) and four reactions involving changes in the promoter state, and each promoter state involves degradation and production reactions. Take the parameter values $\alpha = 1{,}000\,\text{s}^{-1}$, $\beta = 5 \times 10^5$, and $\kappa = 5 \times 10^{-5}\,\text{s}^{-1}$ and consider the two cases (i) $\gamma = 1\,\text{s}^{-1}$ (monostable) and (ii) $\gamma = 0.75\,\text{s}^{-1}$ (bistable). Plot sample trajectories over a time interval of length $T = 10\,\text{min}$ and histogram $m(T)$.

(c) The circadian clock model with stochiometry and propensities listed in Sect. 5.1. Use the following parameter values taken from [228]: $k = 0.5\,\text{nM}\,\text{h}^{-1}, \gamma = 0.3\,\text{nM}\,\text{h}^{-1}, K_m = 2.0\,\text{nM}, K_m' = 0.2\,\text{nM}$, $r = 2.0\,\text{h}^{-1}, \gamma_p = 1.5\,\text{nM}\,\text{h}^{-1}, K_p = 0.1\,\text{nM}$, and $k_1 = k_2 = 0.2\,\text{h}^{-1}$. Plot a sample trajectory of the number of mRNA M and the number of cytosolic clock proteins X_C as a function of time, and check that the oscillation period is around 22 h. Compare with solutions of the deterministic kinetic rate equations. Also plot several sample trajectories in the (M, X_C) phase plane superimposed on the deterministic limit cycle.

Problem 6.12 (Keizer's paradox.). Consider the following autocatalytic reaction scheme:

$$X \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} 2X, \quad X \xrightarrow{k_2} \emptyset.$$

(a) Write down the deterministic kinetic equation for the concentration x of the chemical species X. Show that it has an unstable fixed point at $x_1 = 0$ and a stable fixed point at $x_2 = (k_1 - k_2)/k_{-1}$.

(b) Construct the master equation for the probability $P(n,t)$ that there are n molecules of X at time t. By writing out the explicit equations for $dP(0,t)/dt$, $dP(1,t)/dt$ etc., use induction to show that the unique steady-state solution $P^*(n)$ is

$$P^*(0) = 1, \quad P^*(n) = 0, \quad n > 0.$$

Hence the stochastic model shows that there will be no X left in the system – $X = 0$ is an absorbing state. This appears to contradict the deterministic limit, which is known as Keizer's paradox.

(c) Use the Gillespie algorithm to explore the evolution of the probability distribution $P(m,t)$ as a function of time. In particular, demonstrate that at intermediate times the distribution localizes around the deterministic steady-state x_2 before eventually forming a peak around zero. Hence, provide an explanation of Keizer's paradox in terms of the non-commutativity of the operations $\lim t \to \infty$ and $\lim \Omega \to \infty$ where Ω is the system size.

Part II
Advanced Topics

Chapter 7
Transport Processes in Cells

The efficient delivery of proteins and other molecular products to their correct location within a cell (intracellular transport) is of fundamental importance to normal cellular function and development [3]. Moreover, the breakdown of intracellular transport is a major contributing factor to many degenerative diseases. Broadly speaking, there are four basic mechanisms for intracellular transport [72] (see Fig. 7.1):

(i) *Passive diffusion* within the cytosol or the surrounding plasma membrane of the cell. Since the aqueous environment (cytosol) of a cell is highly viscous at the length and velocity scales of macromolecules (low Reynolds number), a diffusing particle can be treated as an overdamped Brownian particle where inertial effects are ignored (Chap. 2).

(ii) *Active motor-driven transport* along polymerized filaments such as microtubules and F-actin that comprise the cytoskeleton [275] (Chap. 4). Newly synthesized products from the nucleus are mainly transported to other intracellular compartments or the cell membrane via a microtubular network that projects radially from organizing centers (centrosomes). The same network is used to transport degraded cell products back to the nucleus and is also exploited by various animal viruses including HIV in order to reach the nucleus from the cell surface and release their genome through nuclear pores. Active transport is faster and more easily regulated than passive diffusion, but requires a constant supply of energy to do useful work.

(iii) *Transport through membrane pores and channels.* This involves a variety of mechanisms, depending on the size of the transported molecules relative to the width of the membrane and whether the transport process is passive or active. (A pore refers to the case of a thin membrane, whereas a channel refers to the case of a thick membrane, i.e., a long pore.) *Facilitated diffusion* is the passive transport of small solute molecules including ions across a biological membrane via specific transmembrane integral proteins such as ion channels. It is passive in the sense that facilitated transport does not involve the use of chemical energy; rather, molecules and ions move down their concentration gradient. However, it is distinct from diffusion, since it would be difficult for molecules

P.C. Bressloff, *Stochastic Processes in Cell Biology*, Interdisciplinary Applied Mathematics 41, DOI 10.1007/978-3-319-08488-6_7

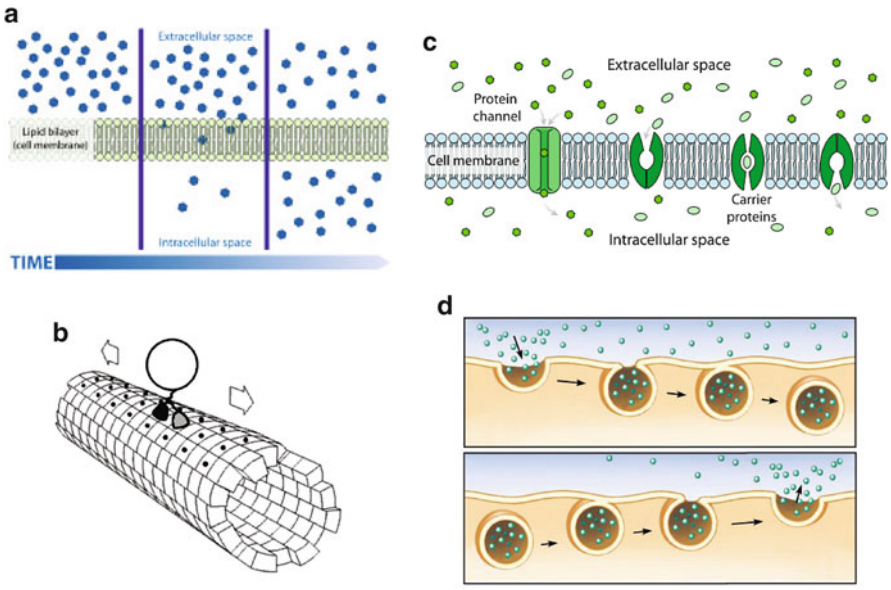


Fig. 7.1: Different mechanisms of intracellular transport: (**a**) Passive diffusion through a cell membrane. (**b**) Active motor-driven transport. (**c**) Facilitated diffusion. (**d**) Exocytosis and endocytosis. (Public domain figures downloaded from Wikipedia Commons.)

or ions to cross the membrane without assistance from the transmembrane proteins. There is also an active form of membrane transport involving ion pumps, which are rotary motors that move ions across a membrane against their concentration gradient.

(iv) *Exocytosis/endocytosis and secretory trafficking.* An important mechanism for regulating the distribution of proteins and lipids in the plasma membrane and intracellular compartments is through endocytosis and exocytosis [246]. *Endocytosis* is the physical process whereby a vesicle forms within a membrane (budding) and is then released into the cytoplasm, whereas *exocytosis* is the complementary process in which an intracellular vesicle fuses with the membrane and releases its contents into a particular compartment or secretes its contents externally (as in the release of neurotransmitters at a synapse). One important role of exocytosis/endocytosis is to regulate protein receptors in the plasma membrane [159, 371] (see Fig. 7.2). That is, molecular motors transport internalized receptors from the plasma membrane to intracellular compartments that either recycle the receptors to the cell surface (*early endosomes* and *recycling endosomes*) or sort them for degradation (*late endosomes* and *lysosomes*). When this is coupled with membrane diffusion, it is possible to regulate the number of receptors within the membrane, which plays an important role in synaptic plasticity. Another important role of exocytosis/endocytosis occurs in the early secretory pathway of eukaryotic cells, which is a critical

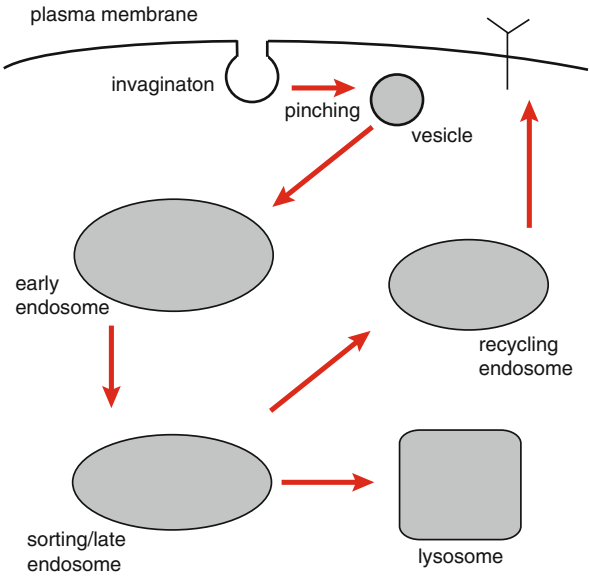


Fig. 7.2: Cartoon sketch of the endocytic pathway. A small region of the plasma membrane is invaginated and pinched off into a vesicle. The internalized vesicle fuses with larger, early endosomes and is then trafficked to sorting endosomes. From there, internalized material either is sent to recycling endosomes followed by reinsertion into the cell surface or is targeted for lysosomal degradation

system for the maturation and transportation of newly synthesized lipids and proteins to specific target sites within the cell membrane [393]. The continuous exchange of proteins and lipid membrane between different compartments of the early secretory pathway will be considered in Sect. 8.4, as an example of a self-organizing process.

Transport within living cells has a number of characteristic features that reflect the complex nature of the cellular environment:

1. The intracellular environment is extremely crowded with macromolecules, subcellular compartments, and confinement domains (Box 1A), suggesting that anomalous subdiffusion is likely to occur [143]. The plasma membrane is also a complex heterogeneous environment [364, 654]. Thus, many papers model diffusion in such environments in terms of CTRWs and fractional Brownian motion (FBM). However, it is still unclear to what extent intracellular diffusion is anomalous in the long-time limit rather than just at intermediate times. This motivates studying diffusion in the presence of obstacles and transient traps whereby normal diffusion is recovered asymptotically [480, 562, 564, 565].
2. Molecules inside the cell are often confined to a domain with small exits on the boundary of the domain. Examples include an ion searching for an open ion channel within the cell membrane, the transport of newly transcribed mRNA

from the nucleus to the cytoplasm via nuclear pores, the confinement of neurotransmitter receptors within a synapse of a neuron, and the confinement of calcium and other signaling molecules within subcellular compartments such as dendritic spines of neurons. This has led to recent interest in using Green's function and asymptotic methods to solve the so-called narrow escape problem [29, 267, 511, 576].

3. A related class of problems involves the search for a small target within the interior of a cellular domain. In this case it is necessary to extend the Smoluchowski theory of diffusion-limited reaction rates [124, 323, 523, 536, 603] to bounded domains or to more complex transport processes than simple diffusion. One example is the arrival of a receptor at a localized reaction site on the surface of an immune cell, which is a key step in the signaling cascade responsible for activating the cell [126]. Another important example is a promoter protein searching for its binding site on DNA, which is facilitated by an intermittent search process in which the particle switches between 3D and 1D diffusion [41, 42, 127, 250, 346, 440, 585].
4. One of the characteristic features of channel diffusion is that it is spatially confined, which leads to strong entropic effects due to the reduction in the available degrees of freedom [83] (see Sect. 1.4 for a definition of entropy). Moreover, various mechanisms of facilitated diffusion can occur through interactions between a diffusing particle and proteins within the channel, as exemplified by nuclear pore complexes (NPCs), which are the sole mediators of exchange between the nucleus and cytoplasm [27]. When a channel becomes sufficiently narrow, particles are no longer able to pass each other (single-file diffusion)—one then finds that a tagged particle exhibits anomalous subdiffusion on long time scales [22]. Entropic effects also play a major role in the *translocation* of biological polymers such as DNA through nanopores.
5. There have been many stochastic models of motor-driven transport at multiple spatial and temporal scales, ranging from Brownian ratchet models [529] to random walk models [391, 457] to systems of PDEs [524, 601]. However, many of these treatments neglect the fact that the goal of such transport is to deliver molecular cargo to specific sites. This then naturally leads to a connection with random intermittent search processes [68, 397, 472, 473]. It also raises the important question regarding signaling mechanisms responsible for localizing a motor complex at a target. Another issue in active transport involves exclusion effects due to multiple motors moving along the same filament track [52, 115, 567].

In this chapter, we present a variety of models of passive and active transport in cells that address various features of the complex cellular environment (see also the review [72]). We begin by considering the anomalous effects of molecular crowding and trapping, where the differences in diffusive behavior at multiple timescales are highlighted (Sect. 7.1). We consider two different mathematical formulations of diffusion–trapping, one based on a reaction–diffusion model and the other based on a CTRW. The particular example of anomalous diffusion in the plasma membrane

is also considered. In Sect. 7.2 we describe how Green's functions (Box 3C) and singular perturbation theory can be used to analyze narrow escape problems and diffusion to a small target. The diffusive transport through narrow membrane pores and channels is described in Sect. 7.3, including the examples of nuclear pores and translocation of DNA. We then turn to PDE models of active motor transport in Sect. 7.4. These focus on the transitions between different types of motion (e.g., anterograde vs. retrograde active transport, diffusion vs. active transport) rather than the microscopic details of how a motor performs a single step. It is shown how such models can be reduced to an effective Fokker–Planck (FP) equation, under the assumption that the transitions between the different motile states are relatively fast—the so-called QSS approximation. In Sect. 7.5 we consider the effects of molecular crowding of motors on a filament track, as modeled by so-called asymmetric exclusion processes (ASEPs). We show how, in the mean-field limit, molecular crowding can be treated in terms of *quasilinear PDEs* that support shock waves [554]. Finally, in Sect. 7.6 the efficiency of various transport processes in targeting a particular subcellular domain is analyzed in terms of the theory of random intermittent search processes. We consider the examples of a transcription factor searching for a promoter site on DNA and the active motor transport of cargo to synaptic targets.

7.1 Anomalous Diffusion

7.1.1 Molecular Crowding, Diffusion-Trapping, and Long-Time Correlations

In normal (unobstructed) diffusion in d dimensions, the MSD of a Brownian particle is proportional to time, $\langle R^2 \rangle = 2dDt$, which is a consequence of the central limit theorem. A general signature of anomalous diffusion is the power law behavior [58, 438] (see Fig. 7.3)

$$\langle R^2 \rangle = 2dDt^{\alpha} \tag{7.1.1}$$

corresponding to either subdiffusion ($\alpha < 1$) or superdiffusion ($\alpha > 1$). Due to recent advances in SPT methods (Sect. 1.2), subdiffusive behavior has been observed for a variety of biomolecules and tracers within living cells. There are a number of subcellular mechanisms thought to generate subdiffusive motion of particles in cells, each with its own distinct type of physical model:

(i) *Molecular crowding*—one of the characteristic features of the interior aqueous environment of cells (cytoplasm) and intracellular compartments such as the endoplasmic reticulum and mitochondria is that they are crowded with small solutes, macromolecules, and skeletal proteins, which occupy 10–50 % of the volume [143]. Cell membranes are also crowded environments containing lipids (molecules consisting of nonpolar, hydrophobic hydrocarbon chains that end in a polar hydrophylic head), which are often organized into raft structures,

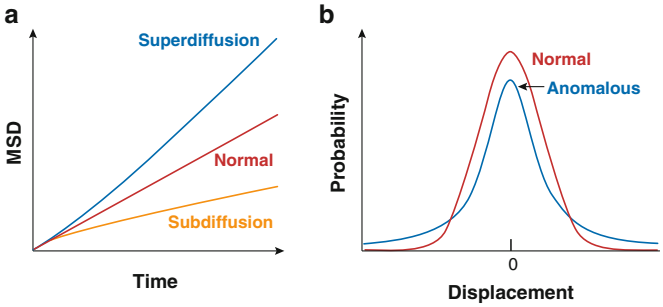


Fig. 7.3: Anomalous diffusion. Two characteristic features of anomalous diffusion are shown: superlinear or sublinear variation of the mean-square displacement (*MSD*) with time; large tails in the probability density

and various mobile and immobile proteins [364]. If the concentration of obstacles is sufficiently high, then subdiffusive behavior occurs, in which the domain of free diffusion develops a fractal-like structure [562]. There is an ongoing debate whether molecular crowding results in anomalous diffusion or leads to a simple reduction in the normal diffusion coefficient on long time scales.

(ii) *Diffusion–trapping*—if a diffusing particle encounters a binding site, then it will pause for a while before dissociating and diffusing away. Multiple binding events with a range of rate constants can generate long tails in the waiting time distribution leading to subdiffusive behavior [564, 565]. Diffusion–trapping is typically modeled in terms of a CTRW [289, 569]; see below.

(iii) *Long-time correlations*—this mechanism involves the viscoelastic properties of the cytoplasm due to the combined effects of macromolecular crowding and the presence of elastic elements such as nucleic acids and cytoskeletal filaments. As a particle moves through the cytoplasm, the latter "pushes back," thus generating long-time correlations in the particle's trajectory. This memory effect can lead to subdiffusive behavior that can be modeled in terms of FBM or the fractional Langevin equation (FLE) [86, 411, 668].

Determining which type of stochastic model best fits experimental data is a nontrivial task, particularly since CTRW, diffusion on fractals, and FBM/FLE generate similar scaling laws for ensemble-averaged behavior in the long-time limit. Thus other measures such as ergodicity (equivalence of time averages and ensemble averages) are being used to help identify which model provides the best characterization for anomalous diffusion in living cells [619, 673, 675]. A more fundamental difficulty in experimentally establishing the existence of anomalous diffusion is that the behavior of $\langle R^2 \rangle$ can depend on the spatial or temporal scale over which observations are made. Consider, for example, the effects of obstacles on protein diffusion [562, 616]. The presence of obstacles reduces the space available for diffusion and consequently decreases the effective diffusion coefficient. As the volume or area fraction of obstacles ϕ is increased, there is a fragmentation of the available space in the sense that many paths taken by a diffusing protein terminate in a dead end

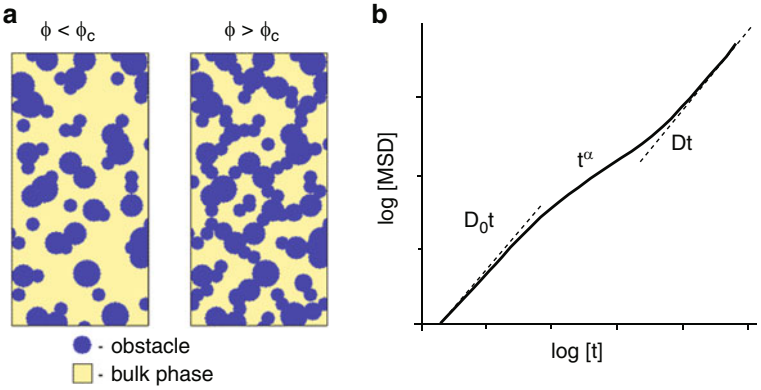


Fig. 7.4: (**a**) Filling of bulk cytoplasm by obstacles with low ($\phi < \phi_c$) and high ($\phi > \phi_c$) volume fraction, where ϕ_c is the percolation threshold. (**b**) Sketch of MSD $\langle R^2 \rangle$ against time t, illustrating three different diffusion regimes below the percolation threshold: unobstructed diffusion (MSD $\sim D_0 t$), anomalous intermediate diffusion MSD $\sim t^\alpha$, and normal effective diffusion MSD $\sim Dt$

and thus do not contribute to diffusive transport (see Fig. 7.4). The region of free diffusion develops a fractal-like structure resulting in anomalous diffusion at intermediate times, $\langle R^2 \rangle \sim t^\alpha$, $\alpha < 1$. (For sufficiently small times $\sqrt{Dt} \ll \xi$, where ξ is the mean distance between obstacles, so that diffusion is normal.) However, assuming that the volume or area fraction is below the percolation threshold (i.e., there is still an unobstructed path across the domain), diffusion is expected to be normal on sufficiently long time scales, $\langle R^2 \rangle \sim t$. On the other hand, above the percolation threshold, proteins are confined and $\langle R^2 \rangle$ saturates as $t \to \infty$. The time it takes to cross over from anomalous to normal diffusion increases with the volume or area fraction ϕ and diverges at the percolation threshold ϕ_c where $\langle R^2 \rangle \sim t^\alpha$ for all times.

Another difficulty in interpreting experimental data is that there are certain practical limitations of current methods [143]. The most effective method for describing membrane diffusion is SPT, as described in Sect. 1.2. This involves the selective labeling of proteins or lipids with fluorophores such as quantum dots, GFP, or organic dyes so that continuous high-resolution tracking of individual molecules can be carried out. SPT can yield nanometer spatial resolution and submillisecond temporal resolution of individual trajectories. However, it is not currently suitable for measuring diffusion in three dimensions due to the relatively rapid speed of 3D diffusion and the problems of imaging in depth. Hence, in the case of diffusion within the cytosol, it is necessary to use a method such as FRAP. Here fluorescently labeled molecules are introduced into the cell and those in some specified volume are bleached by a brief intense laser pulse. The diffusion of unbleached molecules into the bleached volume is then measured. FRAP is limited because it only provides ensemble averaged information of many fluorescent particles, and it also has a restricted measurement time, making it difficult to capture long-tail phenomena expected in anomalous subdiffusion.

7.1.2 Diffusion–Trapping Model of Protein Trafficking in Dendrites

In the Smoluchowski theory of reaction kinetics (Sect. 2.4), it is assumed that when a diffusing particle reacts with the target it disappears, that is, we have the trapping reaction $A + B \to B$ where A denotes a diffusing particle and B denotes an immobile trap. However, within the context of intracellular transport, there are many examples where there is transient trapping of diffusing particles, resulting in anomalous diffusion on intermediate time scales and normal diffusion on long time scales. This has been elucidated by Saxton [564, 565], who carried out Monte Carlo simulations of random walks on a 2D lattice with a finite hierarchy of binding sites, that is, binding sites with a finite set of energy levels. This means that there are no traps that have an infinite escape time so that diffusing particles ultimately equilibrate with the traps and diffusion becomes normal. On the other hand, in the case of infinite hierarchies, arbitrarily deep traps exist but are very rare, resulting in a nonequilibrium system in which anomalous subdiffusion occurs at all times [58]. The latter process can be modeled in terms of a CTRW (see Sect. 7.1.3). Here we will consider a particular example of diffusive transport in the presence of transient immobile traps, namely, the effects of dendritic spines on diffusive protein transport in the dendrites of neurons.

Neurons are among the largest and most complex cells in biology (see Fig. 3.11 of Sect. 3.5). Their intricate geometry presents many challenges for cell function, in particular with regard to the efficient delivery of newly synthesized proteins from the cell body or soma to distant locations on the axon or dendrites. The axon contains ion channels for action potential propagation and presynaptic active zones for neurotransmitter release, whereas each dendrite contains postsynaptic domains (or densities) where receptors that bind neurotransmitter tend to cluster. At most excitatory synapses in the brain, the postsynaptic density (PSD) is located within a dendritic spine, which is a small, sub-micrometer membranous extrusion that protrudes from a dendrite [604] (see Fig. 7.5). Typically spines have a bulbous head that is connected to the parent dendrite through a thin spine neck, and there can exist thousands of spines distributed along a single dendrite. It is widely thought that spines act to compartmentalize chemical signals generated by synaptic activity, thus impeding their diffusion into dendrites [550, 698]. Conversely, in the case of signaling molecules diffusing along the dendrite, the spines act as transient traps as illustrated in Fig. 7.6a. Following along similar arguments to the case of diffusion in the presence of obstacles, normal diffusion is expected at short and long times and anomalous subdiffusion at intermediate times. Anomalous subdiffusion was indeed observed experimentally by Santamaria et al. [556], such that the MSD $\langle \mathbf{x}^2(t) \rangle \sim D_0 t^{2/\beta}$ at intermediate times with $\beta > 2$ and D_0 the free diffusion coefficient. As might be expected, β increases (slower diffusion) with increasing spine density. β also increases when the volume of the spine head is increased relative to the spine neck, reflecting the fact there is an enhanced bottleneck. Note that anomalous diffusion can occur at all times if the reactions within each spine are taken to have a non-exponential waiting time density [181] (see also Sect. 7.1.3).

A related problem is the diffusive transport of neurotransmitter protein receptors within the plasma membrane of a dendrite, with each spine acting as a transient trap that localizes the receptors at a synapse. The majority of fast excitatory synaptic transmission in the central nervous system is mediated by AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole-propionic acid) receptors, which respond to the neurotransmitter glutamate. There is now a large body of experimental evidence that the fast trafficking of AMPA receptors into and out of spines is a major contributor to activity-dependent, long-lasting changes in synaptic strength [63, 123, 257]. SPT experiments (Sect. 1.2 and Fig. 1.5) suggest that surface AMPA receptors diffuse freely within the dendritic membrane until they enter a spine, where they are temporarily confined by the geometry of the spine and through interactions with scaffolding proteins and cytoskeletal elements [111, 160, 213, 642] (see Fig. 7.6b). A surface receptor may also be internalized via endocytosis and stored within an intracellular compartment, where it is either recycled to the surface via recycling endosomes and exocytosis, or sorted for degradation by late endosomes and lysosomes [159] (see Fig. 7.2). A number of single spine models have explored the combined effects of diffusion, trapping, receptor clustering, and recycling on the number of synaptic AMPA receptors [85, 132, 154, 269, 587]. In such models, the synapse

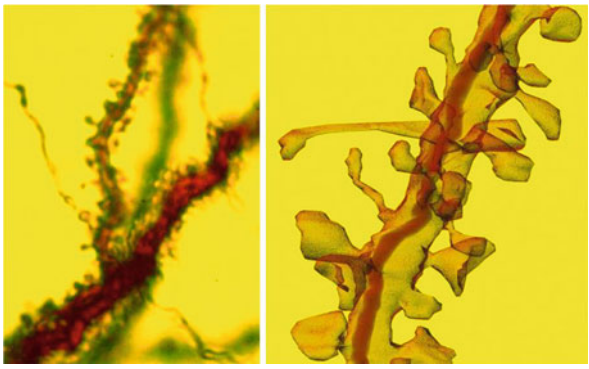


Fig. 7.5: An example of a piece of spine studded dendritic tissue (from rat hippocampal region CA1 stratum radiatum). The dendrite on the right-hand side is $\sim 5\,\mu$m in length. Taken with permission from SynapseWeb, Kristen M. Harris, PI, http://synapses.clm.utexas.edu/

is treated as a self-organizing compartment in which the number of AMPA receptors is a dynamic steady state that determines the strength of the synapse; activity-dependent changes in the strength of the synapse then correspond to shifts in the dynamical set point. When receptor–receptor interactions are included a synapse can exhibit bistability between a non-clustered and clustered state [587], which can be understood in terms of a liquid-vapor phase transition [85].

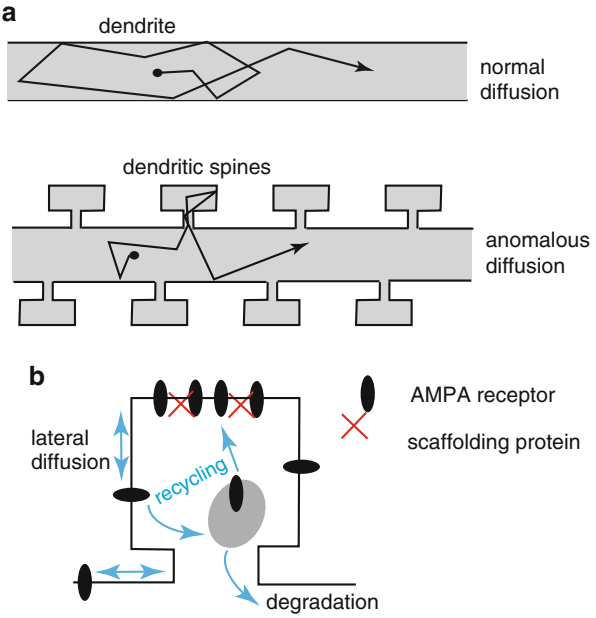


Fig. 7.6: (**a**) Schematic illustration of the anomalous diffusion model of [556], who carried out detailed 3D simulations of diffusion in a spiny dendrite treated as a system of connected cylinders with the following baseline parameter values: spine neck diameter 0.2 µm, neck length 0.6 µm, head length and diameter 0.6 µm, dendrite diameter 1 µm, and a spine density of 15 spines/µm. The dendritic spines act as transient traps for a diffusing particle within the dendrite, which leads to anomalous diffusion on intermediate time scales. (**b**) Schematic illustration of various pathways of AMPA receptor trafficking at a dendritic spine, including lateral diffusion, binding with scaffolding proteins, and recycling between the surface and intracellular compartments

Given the tubular-like structure of a dendrite, it is possible to model the diffusion of proteins in the plasma membrane using a reduced 1D model (analogous to the cable equation for voltage changes along a dendrite [322]). Since the Green's function of the 1D diffusion equation is non-singular (Box 2C), one can treat the dendritic spines as point-like sources or sinks [67, 76]. Consider a population of N identical spines distributed along a uniform dendritic cable of length L and circumference l, with x_j, $j = 1,\ldots,N$, the position (axial coordinate) of the jth spine. Let $p(x,t)$ denote the probability density (per unit area) that a surface receptor is located within the dendritic membrane at position x at time t. Similarly, let $R_j(t)$ denote the probability that the receptor is trapped at the surface of the jth spine. A simple 1D diffusion-trapping model of AMPA receptor trafficking takes the form [67, 155]

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2} - \sum_{j=1}^{N} h[p_j - R_j/A]\delta(x - x_j), \tag{7.1.2a}$$

$$\frac{dR_j}{dt} = lh[p_j - R_j/A], \tag{7.1.2b}$$

where D is the diffusivity in the plasma membrane, A is the surface area of a spine, and $p_j(t) = p(x_j,t)$. Equation (7.1.2a) needs to be supplemented by boundary conditions at the ends of the cable. Suppose that there exists a steady-state solution, which occurs if there are reflecting boundary conditions at either end. Integrating the right-hand side of Eq. (7.1.2a) over the interval $[x_j - \Delta, x_j + \Delta]$, where Δ is the effective half-width of a spine, and multiplying by the circumference l leads to the current conservation condition

$$l[J(x_j - \Delta) - J(x_j + \Delta)] = lh[p_j - R_j/A], \quad J(x) = -D\partial_x p.$$

Hence, $lh[p_j - R_j/A]$ is the probability per unit time that a receptor enters the spine with h an effective hopping rate. (This rate depends on the detailed geometry of the dendritic spine.)

The effective diffusivity of a receptor in the long-time limit, which takes into account the effects of trapping at spines, can be determined by calculating the MFPT $\tau(X)$ to travel a distance X from the end $x = 0$, for example. Introducing an absorbing boundary at $x = X$ and a reflecting boundary at $x = 0$,

$$Dl\partial_x p(0,t) = -\delta(t), \quad p(X,t) = 0,$$

the function

$$\mathbb{P}(X,t) \equiv l\int_0^X p(x,t)dx + \sum_{j=1}^{N_X} R_j(t) \tag{7.1.3}$$

is then the probability that $t < \tau(X)$, i.e., the probability that a receptor which was initially at the origin has not yet reached the point $x = X$ in a time t. Here N_X is the number of spines in the interval $[0,X)$. The MFPT is then $\tau(X) = \int_0^\infty \mathbb{P}(X,t)dt$. It follows that the MFPT can be expressed in terms of Laplace transforms:

$$\tau(X) = \int_0^X \tilde{p}(x,0)dx + \sum_{j=1}^{N_X} \tilde{R}_j(0) \tag{7.1.4}$$

where $\tilde{f}(s) \equiv \int_0^\infty \mathrm{e}^{-st} f(t)dt$ (see Box 2A). Laplace transforming Eqs. (7.1.2a) and (7.1.2b) and using the initial conditions $p(x,0) = 0$ for $0 < x < X$ and $R(0) = 0$ gives

$$-s\tilde{p} + D\frac{\partial^2 \tilde{p}}{\partial x^2} = \sum_{j=1}^{N_X} h[\tilde{p}_j - \tilde{R}_j/A]\delta(x - x_j), \tag{7.1.5a}$$

$$s\tilde{R}_j = lh[\tilde{p}_j - \tilde{R}_j/A], \tag{7.1.5b}$$

where $\tilde{p}_j(s) = \tilde{p}(x_j, s)$. In the limit $s \to 0$, Eq. (7.1.5b) implies that $A\tilde{p}_j(0) = \tilde{R}_j(0)$, and Eq. (7.1.5a) becomes

$$D\frac{\partial^2 \tilde{p}(x,0)}{\partial x^2} = 0. \tag{7.1.6}$$

Imposing the boundary conditions at $x = 0, X$, which imply that $l\partial_x \tilde{p}(x,0)|_{x=0} = -1/D$ and $\tilde{p}(X,0) = 0$, we find that $l\tilde{p}(x,0) = (X-x)/D$. Combining these results,

$$\tau(X) = \frac{X^2}{2D} + \frac{A}{lD}\sum_{j=1}^{N_X}(X - x_j). \tag{7.1.7}$$

The first term on the right-hand side of this equation is the MFPT in the absence of any spines, whereas the remaining terms take into account the effects of being temporarily trapped at a spine.

In order to calculate an effective diffusivity, consider the simple example of identical spines distributing uniformly along the cable with spacing d. That is, $x_j = jd$, $j = 1, \ldots, N$ such that $Nd = L$ and $N_X = X/d$ for $X \gg d$. Equation (7.1.7) then becomes (for $N_X \gg 1$)

$$\tau(X) \equiv \frac{X^2}{2D_{\rm eff}} = \frac{X^2}{2D} + \frac{A}{lD}\sum_{j=1}^{N_X}(X - jd).$$

Using the approximation

$$\sum_{j=1}^{N_X}(X - jd) = N_X X - \frac{(N_X+1)N_X d}{2} \approx \frac{X^2}{2d}$$

finally gives [67]

$$D_{\rm eff} = \frac{D}{1 + A/ld}. \tag{7.1.8}$$

As expected, the presence of traps reduces the effective diffusivity of a receptor. In particular, the diffusivity is reduced by increasing the surface area A of a spine relative to the product of the spine spacing d and circumference l of the dendritic cable. Interestingly, D does not depend on the hopping rate h. Taking typical measured values of the area of a spine ($A = 1\,\mu\text{m}^2$), the spacing between spines ($d = 1\,\mu\text{m}$), and the circumference of a dendrite ($l = 1\,\mu\text{m}$) [604], it follows that $D_{eff} = 0.5D$. A much greater reduction in the effective diffusivity can be obtained by including additional aspects of receptor dynamics in spines such as binding/unbinding to cytoskeletal proteins (see Ex. 7.1) and recycling between the surface and intracellular compartments [67, 76]. In the latter case, Eqs. (7.1.2a) and (7.1.2b) become

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2} - \sum_{j=1}^{N} h[p_j - R_j/A]\delta(x - x_j), \tag{7.1.9a}$$

$$\frac{dR_j}{dt} = lh[p_j - R_j/A] - \sigma_{\text{end}}R_j + \sigma_{\text{exo}}S_j \tag{7.1.9b}$$

$$\frac{dS_j}{dt} = -\sigma_{\text{exo}}S_j + \sigma_{\text{end}}R_j, \tag{7.1.9c}$$

where $S_j(t)$ is the probability that a receptor is in an intracellular compartment associated with the jth spine and σ_{exo} (σ_{end}) is the rate of exocytosis (endocytosis) between the compartment and the surface of the spine. For simplicity, we neglect degradation of receptors by lysosomes. The analysis of this model proceeds in an identical fashion to the binding model considered in Ex. 7.1. It is also possible to use diffusion–trapping models to determine the steady-state distribution of receptors when there is a nonzero flux of receptors at one end (representing newly synthesized receptors from the cell body) and the degradation of receptors is included (see Ex. 7.2).

One major simplification of the diffusion-trapping model is that it neglects the detailed structure of a spine and the associated PSD. A more comprehensive model would need to take into account the complex organization of the PSD and the geometry of the spine [194, 330, 406, 580]. Finally, note that the coupling between exocytosis and endocytosis during AMPA receptor recycling is one example of a more general transport mechanism that occurs in neurons and other secretory cells via the endocytic pathway illustrated in Fig. 7.2 [246]. Other examples include the insertion and removal of membrane proteins during axonal elongation and the stimulus-induced release of secretory molecules (neurotransmitters) at the presynaptic terminal of a synapse. The latter is regulated by the exocytosis of synaptic vesicles; endocytic processes then have to be coordinated so that there is an efficient reuptake of vesicles in order to restore functionality of the synapse. For a detailed discussion of whole-cell kinetic models of receptor recycling and its role in chemical signaling see [371, 680].

7.1.3 Continuous-Time Random Walks

An important generalization of the standard random walk, known as the CTRW, is often used to model anomalous diffusion arising from trapping processes. The basic idea is that trapping can increase the time between jumps (waiting times) so that jumps no longer occur at fixed discrete-time steps. A CTRW is typically written in the form [289]

$$R_n(\ell,t) = \sum_{\ell' \in \Gamma} p(\ell - \ell') \int_0^t \psi(t - t')R_{n-1}(\ell',t')dt' \tag{7.1.10}$$

where $R_n(\ell,t)$ is the probability density for a walker to just arrive at site ℓ at time t in n steps and $\psi(t)$ is the waiting time density for a single step over a time interval of length t. Thus steps can now take place at different times. In general, a CTRW has a memory of previous time steps so it is non-Markovian. How does one recover the standard Markovian random walk? First, we consider the exponential waiting time density $\psi(t)$ given by

$$\psi(t) = \Lambda \mathrm{e}^{-\Lambda t}.$$

Substituting into the CTRW and differentiating both sides with respect to t shows that

$$\frac{1}{\Lambda}\frac{dR_n}{dt} + R_n(\ell,t) \approx R_n(\ell,t+\varepsilon) = \sum_{\ell'} p(\ell-\ell')R_{n-1}(\ell',t)$$

for $\varepsilon = \Lambda^{-1} \ll 1$. This is the standard recursive equation after setting $t = n\varepsilon$.

In order to analyze Eq. (7.1.10) we will make extensive use of the transform methods described in Box 2A. It is useful to keep track of the different transform pairs, which are as follows:

$$n \leftrightarrow z\,(z-\text{transform}),\ \ell \leftrightarrow \mathbf{k}\,(\text{discrete Fourier transform}),\ t \leftrightarrow s\,(\text{Laplace transform}).$$

First, taking the Laplace transform of Eq. (7.1.10) and using the convolution Theorem 2.1 give

$$\tilde{R}_n(\ell,s) = \tilde{\psi}(s)\sum_{\ell'\in\Gamma} p(\ell-\ell')\tilde{R}_{n-1}(\ell',s).$$

This has the solution (for $n \geq 1$)

$$\tilde{R}_n(\ell,s) = AP_n(\ell)\tilde{\psi}_n(s), \quad \tilde{\psi}_n(s) = \tilde{\psi}(s)^n$$

with unknown amplitude A and $P_n(\ell)$ the solution to the standard random walk (RW) master equation

$$P_n(\ell) = \sum_{\ell'} p(\ell-\ell')P_{n-1}(\ell').$$

One is often interested in the probability density of being at the site ℓ at time t irrespective of the number of steps. Therefore, introduce the density

$$R(\ell,t) = \sum_{n\geq 0} R_n(\ell,t), \tag{7.1.11}$$

and note that if the walker starts at the origin, then

$$R_0(\ell,t) = \delta_{\ell,0}\Psi(t), \quad \Psi(t) = \int_t^\infty \psi(t')dt'.$$

Here $\Psi(t)$ is the probability that the walker has not yet taken a first step at time t. It follows from the analysis of $R_n(\ell,t)$ and its Laplace transform that

$$R(\ell,t) = \sum_{\ell'\in\Gamma} p(\ell-\ell')\int_0^t \psi(t-t')R(\ell',t')dt' + \delta_{\ell,0}\Psi(t) \tag{7.1.12}$$

and

$$\tilde{R}(\boldsymbol{\ell},s)=\sum_{n\geq 0}\tilde{R}_n(\boldsymbol{\ell},s)=A\sum_{n\geq 0}P_n(\boldsymbol{\ell})\tilde{\psi}(s)^n.$$

Summing both sides with respect to $\boldsymbol{\ell}$ and using the normalization $\sum_{\boldsymbol{\ell}}P_n(\boldsymbol{\ell})=1$, we have

$$\sum_{\boldsymbol{\ell}}\tilde{R}(\boldsymbol{\ell},s)=A\sum_{n\geq 0}\tilde{\psi}(s)^n=\frac{A}{1-\tilde{\psi}(s)}.$$

We also have $\sum_{\boldsymbol{\ell}}R(\boldsymbol{\ell},t)=1$, which implies $\sum_{\boldsymbol{\ell}}\tilde{R}(\boldsymbol{\ell},s)=1/s$. Hence, $A=(1-\tilde{\psi}(s))/s$ and

$$\tilde{R}(\boldsymbol{\ell},s)=\frac{1-\tilde{\psi}(s)}{s}\Gamma[\boldsymbol{\ell},\tilde{\psi}(s)], \tag{7.1.13}$$

where $\Gamma(\boldsymbol{\ell},z)=\sum_n P_n(\boldsymbol{\ell})z^n$, which is the generating function for the distribution $P_n(\boldsymbol{\ell})$ and fixed $\boldsymbol{\ell}$.

We would like to determine the MSD for different choices of the waiting time density $\psi(t)$. First, we need to determine the generating function Γ for the underlying RW, which satisfies

$$\Gamma(\boldsymbol{\ell},z)=\delta_{\boldsymbol{\ell},0}+z\sum_{\boldsymbol{\ell}'}p(\boldsymbol{\ell}-\boldsymbol{\ell}')\Gamma(\boldsymbol{\ell}',z).$$

Taking discrete Fourier transforms and using the convolution theorem on the (infinite) lattice

$$\hat{\Gamma}(\mathbf{k},z)=1+z\hat{p}(\mathbf{k})\hat{\Gamma}(\mathbf{k},z),$$

where

$$\hat{p}(\mathbf{k})=\sum_{\boldsymbol{\ell}}\mathrm{e}^{i\mathbf{k}\cdot\boldsymbol{\ell}}p(\boldsymbol{\ell}).$$

Hence

$$\hat{\Gamma}(\mathbf{k},z)=\frac{1}{1-z\hat{p}(\mathbf{k})}.$$

Taking the discrete Fourier transform of Eq. (7.1.13) with $\widehat{R}(\mathbf{k},s)\equiv\sum_{\boldsymbol{\ell}}\mathrm{e}^{i\mathbf{k}\cdot\boldsymbol{\ell}}\tilde{R}(\boldsymbol{\ell},s)$ thus gives

$$\widehat{R}(\mathbf{k},s)=\frac{1-\tilde{\psi}(s)}{s}\frac{1}{1-\tilde{\psi}(s)\hat{p}(\mathbf{k})}. \tag{7.1.14}$$

For simplicity, consider a 1D lattice. Using

$$\left(-i\frac{\partial}{\partial k}\right)^n\widehat{R}(k,s)\bigg|_{k=0}=\sum_{\ell}\ell^n\tilde{R}(\ell,s),$$

which is the Laplace transform of the nth order moment of ℓ, we obtain the following results:

$$\mathscr{L}(\langle X\rangle)(s)=\sum_{\ell}\ell R(\ell,s)=\frac{-i\tilde{\psi}(s)\hat{p}'(0)}{s(1-\tilde{\psi}(s))}$$

and

$$\mathscr{L}(\langle X^2\rangle)(s) = \sum_\ell \ell^2 R(\ell, s) = \frac{\tilde{\psi}(s)\hat{p}''(0)}{s(1-\tilde{\psi}(s))} - \frac{2\tilde{\psi}(s)^2\hat{p}'(0)^2}{s(1-\tilde{\psi}(s))^2}.$$

We are assuming that the distribution $p(\ell)$ of displacements on the lattice has finite moments. In particular, if the CTRW is unbiased with $\hat{p}'(0) = 0$ and $\hat{p}''(0) = -\sigma^2$, then

$$\mathscr{L}(\langle X\rangle)(s) = 0$$

and

$$\mathscr{L}(\langle X^2\rangle)(s) = \frac{\tilde{\psi}(s)\sigma^2}{s(1-\tilde{\psi}(s))}.$$

Now suppose that the mean waiting time density is also finite:

$$\tau = \int_0^\infty t\,\psi(t)dt < \infty.$$

We can then Taylor expand the Laplace transform around $u = 0$

$$\tilde{\psi}(s) = \int_0^\infty \mathrm{e}^{-st}\psi(t)dt \approx \int_0^\infty [1-st]\psi(t)dt = 1 - s\tau,$$

so that in the limit $s \to 0$,

$$\mathscr{L}(\langle X^2\rangle)(s) \sim \frac{\sigma^2}{s^2\tau},$$

which implies that

$$\langle X(t)^2\rangle \sim \frac{\sigma^2 t}{\tau}.$$

On the other hand, if

$$\tilde{\psi}(s) \sim 1 - Bs^\beta$$

as $s \to 0$ with $0 < \beta < 1$, then the waiting time density has infinite mean and

$$\psi(t) \sim t^{-1-\beta}$$

as $t \to \infty$. One now obtains subdiffusive behavior with

$$\langle X(t)^2\rangle \sim \frac{\sigma^2 t^\beta}{B\Gamma(1+\beta)}. \tag{7.1.15}$$

The relationship between the power laws with respect to t and the Laplace variable u are determined using Tauberian theorems (see Box 7A).

Box 7A. Tauberian theorems [289].

Weak Tauberian theorem. If $\psi(t) \geq 0$, $0 \leq \rho < \infty$ and L is slowly varying at ∞, then each of the relations

$$\tilde{\psi}(s) \sim L(1/s)s^{-\rho}$$

as $u \to 0$ and

$$\int_0^t \psi(t')dt' \sim \frac{t^\rho L(t)}{\Gamma(1+\rho)}$$

as $t \to \infty$ implies the other.

Strong Tauberian theorem. If $\psi(t) \geq 0$, $0 \leq \rho < \infty$, $\psi(t)$ is ultimately monotonic as $t \to \infty$, and L is slowly varying at ∞, then each of the relations

$$\tilde{\psi}(s) \sim L(1/s)s^{-\rho}$$

as $u \to 0$ and

$$\psi(t) \sim \frac{t^{\rho-1}L(t)}{\Gamma(\rho)}$$

as $t \to \infty$ implies the other.

7.1.4 Diffusion in the Plasma Membrane

At the simplest level, the plasma membrane can be treated as a 2D sheet of membrane lipids into which proteins are embedded (see Fig. 7.7). *Membrane lipids* are a group of compounds (structurally similar to fats and oils) which form the double-layered surface of all cells. The three major classes of membrane lipids are phospholipids, glycolipids, and cholesterol. Lipids are amphiphilic: they have one end that is soluble in water ("polar") and an ending that is soluble in fat ("nonpolar"). By forming a double layer with the polar ends pointing outward and the nonpolar ends pointing inward membrane lipids can form a "lipid bilayer" which keeps the watery interior of the cell separate from the watery exterior. The arrangements of lipids and various proteins, acting as receptors and channel pores in the membrane, control the entry and exit of other molecules as part of the cell's metabolism. In the *fluid mosaic model* of [593], the membrane lipids are treated as the solvent (water concentrations are very low within the membrane) into which proteins are dissolved. One of the consequences of the fluid mosaic model is that protein clustering, which alters the effective size of a diffusing particle, has only a weak effect on diffusion in the plasma membrane. This follows from the hydrodynamic membrane diffusion model of Saffman and Delbruck [551], which implies that the diffusion coefficient

for a cylinder of radius r in a 2D membrane varies as $\log r$. Although the diffusion of lipids appears to be Brownian in pure lipid bilayers, SPT experiments (Sect. 1.2) indicate that lipids and proteins undergo anomalous diffusion in the plasma membrane [364, 565]. This has led to a modification of the original fluid mosaic model, whereby lipids and transmembrane proteins undergo confined diffusion within, and hopping between, membrane microdomains or corrals [364, 365, 654]; the corraling could be due to "fencing" by the actin cytoskeleton or confinement by anchored protein "pickets" (see Fig. 7.8). These microdomains could also be associated with lipid rafts [299, 365]. Partitioning the membrane into a set of corrals implies that anomalous diffusion of proteins will be observed on intermediate timescales, due to the combined effects of confinement and binding to the actin cytoskeleton. However, on time scales over which multiple hopping events occur, normal diffusion will be recovered. A rough estimate of the corresponding diffusion coefficient is $D \sim L^2/\tau$, where L is the average size of a microdomain and τ is the mean hopping rate between microdomains. A typical range of values for various types of mammalian cell are $L \sim 30$–$240\,\text{nm}$ and $\tau \sim 1$–$20\,\text{ms}$.

In the case of confinement by anchored protein pickets, τ can be estimated by treating each corral as a domain with a set of small holes (gaps) between anchored proteins and solving a narrow escape problem [267, 269] (see Sect. 7.2). (Another approach to estimating τ has been developed by [327], based on a random walker moving on a 1D lattice with either periodically or randomly distributed semipermeable barriers.) On the other hand, the membrane cytoskeleton surrounding a corral is usually modeled as an effective energy barrier over which a diffusing protein must escape. For example, Saxton carried out a computational study of a particle diffusing inside a corral surrounded by a static energy barrier [563]. It was assumed that when the particle hit the barrier it had a fixed probability of escape. The MFPT out of the corral was numerically determined for a wide range of corral sizes, shapes, and escape probabilities. In earlier work, Saxton considered a static fence model in

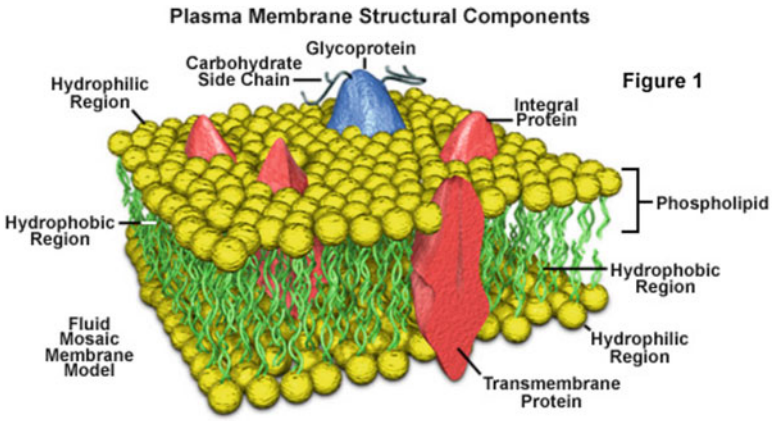


Fig. 7.7: Fluid mosaic model of the plasma membrane. (Public domain figure downloaded from Wikispaces.)

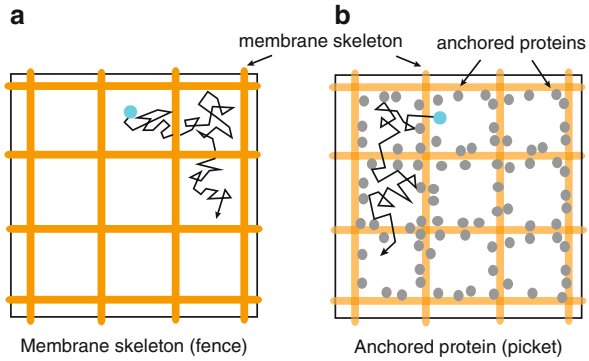


Fig. 7.8: Picket-fence model of membrane diffusion. The plasma membrane is parceled up into compartments whereby both transmembrane proteins and lipids undergo short-term confined diffusion within a compartment and long-term hop diffusion between compartments. This corralling is assumed to occur by two mechanisms. (**a**) The membrane-cytoskeleton (*fence*) model: transmembrane proteins are confined within the mesh of the actin-based membrane skeleton. (**b**) The anchored-protein (*picket*) model: transmembrane proteins, anchored to the actin-based cytoskeleton, effectively act as rows of pickets along the actin fences

which a protein could only move from one corral to another if the particular barrier separating the two corrals was dissociated [561]. In this particular model, large-scale diffusion only occurs if there exists a percolation network of dissociated barriers. However, estimates of the density of the actin cytoskeleton in red blood cells (erythrocytes), for example, suggest that the fraction of dissociated cytoskeleton is below the percolation threshold. Hence, it is necessary to modify the percolation model by considering time-dependent, fluctuating energy barriers. A simple model of this is the stochastic gating model described in Sect. 3.6.

7.2 Narrow Escape Problems, Small Targets, and Singular Perturbation Methods

7.2.1 Narrow Escape Problems

Within the context of intracellular transport, there has been a growing interest in a particular class of FPT processes, namely, the escape of a freely diffusing molecule from a 2D or 3D bounded domain through small absorbing windows on an otherwise reflecting boundary [29, 76, 107, 267, 511, 576]. Examples include an ion searching for an open ion channel situated within the cell membrane, the confinement of neurotransmitter receptors within the synapse of a neuron, and the confinement of calcium within intracellular compartments. Consider diffusion in a two-dimensional domain $\Omega \subset \mathbb{R}^2$ whose boundary can be decomposed as $\partial\Omega = \partial\Omega_r \cup \partial\Omega_a$, where $\partial\Omega_r$ represents the reflecting part of the boundary and $\partial\Omega_a$ the absorbing part. We then have a narrow escape problem in the limit that the measure of the absorbing

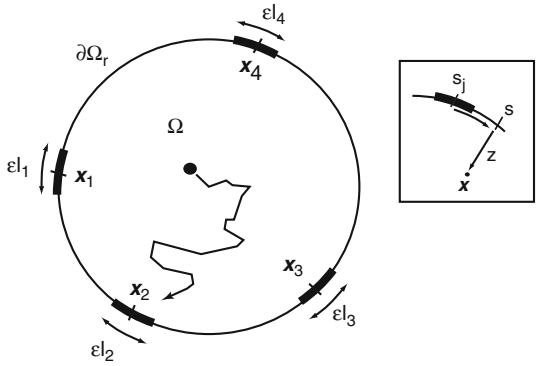


Fig. 7.9: Example trajectory of a Brownian particle moving in a 2D unit disk with small absorbing windows on an otherwise reflecting circular boundary. *Inset*: a local coordinate system around the jth arc

set $|\partial\Omega_a| = \mathscr{O}(\varepsilon)$ is asymptotically small, that is, $0 < \varepsilon \ll 1$. It follows from the analysis of exit times [see Eq. (2.3.8)], that the MFPT to exit the boundary $\partial\Omega_a$ satisfies the equation (in the absence of external forces)

$$\nabla^2 \tau(\mathbf{x}) = -\frac{1}{D}, \quad \mathbf{x} \in \Omega \tag{7.2.1}$$

with boundary conditions

$$\tau(\mathbf{x}) = 0, \quad \mathbf{x} \in \partial\Omega_a = \cup_{j=1}^{N} \partial\Omega_j$$

and

$$\partial_n \tau(\mathbf{x}) = 0, \quad \mathbf{x} \in \partial\Omega_r.$$

The absorbing set is assumed to consist of N small disjoint absorbing windows $\partial\Omega_j$ centered at $\mathbf{x}_j \in \partial\Omega$. In the 2D case, each window is a small absorbing arc of length $|\partial\Omega_j| = \varepsilon l_j$ with $l_j = \mathscr{O}(1)$. It is also assumed that the windows are well separated, that is, $|\mathbf{x}_i - \mathbf{x}_j| = \mathscr{O}(1)$ for all $i \neq j$. An example of a Brownian particle in a 2D unit disk with small absorbing windows on the circular boundary is illustrated in Fig. 7.9. Since the MFPT diverges as $\varepsilon \to 0$, the calculation of $\tau(\mathbf{x})$ requires solving a singular perturbation problem [29, 107, 267, 511, 576].

We follow the particular approach to the narrow escape problem by Ward and collaborators [107, 511]. For a complementary Green's function approach, see the work of Holcman, Schuss, and collaborators reviewed in [268]. The basic idea is to construct the asymptotic solution for the MFPT in the limit $\varepsilon \to 0$ using the method of matched asymptotic expansions. That is, an inner or local solution valid in a $\mathscr{O}(\varepsilon)$ neighborhood of each absorbing arc is constructed and then these are matched to an outer or global solution that is valid away from each neighborhood (see Fig. 7.10). In order to construct an inner solution near the jth absorbing arc, Eq. (7.2.1) is rewritten in terms of a local orthogonal coordinate system (z, s), in

which s denotes arc length along $\partial\Omega$ and z is the minimal distance from $\partial\Omega$ to an interior point $\mathbf{x} \in \Omega$, as shown in the inset of Fig. 7.9. Now introduce stretched coordinates $\hat{z} = z/\varepsilon$ and $\hat{s} = (s - s_j)/\varepsilon$, and write the solution to the inner problem as $\tau(\mathbf{x}) = w(\hat{z}, \hat{s})$. Neglecting terms of $\mathcal{O}(\varepsilon)$, it can be shown that w satisfies the homogeneous equation [511]

$$\frac{\partial^2 w}{\partial^2 \hat{z}} + \frac{\partial^2 w}{\partial^2 \hat{s}} = 0, \quad 0 < \hat{z} < \infty, \quad -\infty < \hat{s} < \infty \tag{7.2.2}$$

with the following boundary conditions on $\hat{z} = 0$:

$$\frac{\partial w}{\partial \hat{z}} = 0 \text{ for } |\hat{s}| > l_j/2, \quad w = 0 \text{ for } |\hat{s}| < l_j/2. \tag{7.2.3}$$

The resulting boundary value problem can be solved by introducing elliptic cylinder coordinates. However, in order to match the outer solution we need only specify the far-field behavior of the inner solution, which takes the form

$$w(\mathbf{x}) \sim A_j \left[\log|\mathbf{y}| - \log d_j + o(1)\right] \text{ as } |\mathbf{y}| \to \infty, \tag{7.2.4}$$

where $d_j = l_j/4$, $|\mathbf{y}| = |\mathbf{x} - \mathbf{x}_j|/\varepsilon = \sqrt{\hat{z}^2 + \hat{s}^2}$, and A_j is an unknown constant, which is determined by matching with the outer solution.

As far as the outer solution is concerned, each absorbing arc shrinks to a point $\mathbf{x}_j \in \partial\Omega$ as $\varepsilon \to 0$ (see Fig. 7.10b). Each point $\mathbf{x}_j$ effectively acts as a point source that generates a logarithmic singularity resulting from the asymptotic matching of the outer solution to the far-field behavior of the inner solution. Thus the outer solution satisfies

$$\nabla^2 \tau(\mathbf{x}) = -\frac{1}{D}, \quad \mathbf{x} \in \Omega, \tag{7.2.5}$$

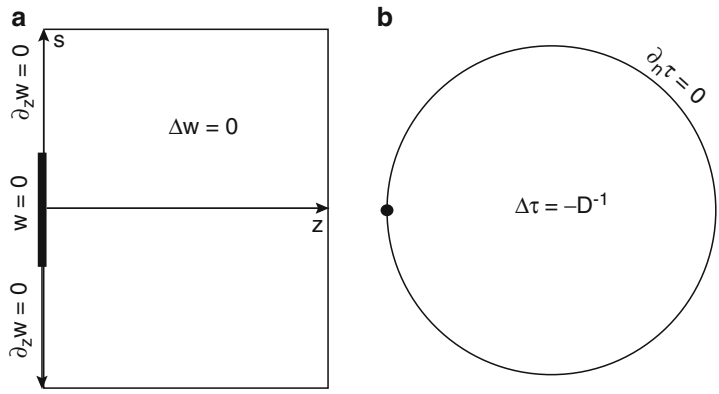


Fig. 7.10: Construction of the matched asymptotic solution for the narrow escape problem. (**a**) Inner solution w in the half-plane $s \in \mathbb{R}, z \in \mathbb{R}^+$ with mixed boundary conditions on $z = 0$. (**b**) Outer solution τ in the disk with a reflecting boundary condition and the target treated as a point

with reflecting boundary condition

$$\partial_n \tau = 0 \text{ for } \mathbf{x} \in \partial\Omega \backslash \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \tag{7.2.6}$$

and

$$\tau(\mathbf{x}) \sim \frac{A_j}{\mu_j} + A_j \log|\mathbf{x} - \mathbf{x}_j| \text{ as } \mathbf{x} \to \mathbf{x}_j,\ j = 1, \ldots, N, \tag{7.2.7}$$

where

$$\mu_j \equiv -\frac{1}{\log(\varepsilon d_j)}. \tag{7.2.8}$$

This can be solved in terms of the Neumann Green's function G, defined as the unique solution of (see Box 2C)

$$\nabla^2 G(\mathbf{x}, \mathbf{x}') = \frac{1}{|\Omega|} - \delta(\mathbf{x} - \mathbf{x}'), \quad x \in \Omega \tag{7.2.9a}$$

$$G(\mathbf{x}, \mathbf{x}_j) \sim -\frac{1}{\pi} \log|\mathbf{x} - \mathbf{x}_j| + R(\mathbf{x}_j, \mathbf{x}_j) \text{ as } \mathbf{x} \to \mathbf{x}_j \in \partial\Omega \tag{7.2.9b}$$

$$\partial_n G(\mathbf{x}, \mathbf{x}') = 0, \mathbf{x} \in \partial\Omega, \quad \int_\Omega G(\mathbf{x}, \mathbf{x}_j) d\mathbf{x} = 0, \tag{7.2.9c}$$

where $R(\mathbf{x}, \mathbf{x}')$ is the regular part of $G(\mathbf{x}, \mathbf{x}')$. It follows that the outer solution can be expressed as

$$\tau(\mathbf{x}) = -\pi \sum_{j=1}^{N} A_i G(\mathbf{x}, \mathbf{x}_j) + \chi, \tag{7.2.10}$$

where χ is an unknown constant. Integrating both sides of Eq. (7.2.10) shows that χ is the MFPT averaged over all possible starting positions:

$$\chi = \overline{\tau} \equiv \frac{1}{|\Omega|} \int_\Omega \tau(\mathbf{x}) d\mathbf{x}. \tag{7.2.11}$$

The problem has reduced to solving $N+1$ linear equations for $N+1$ unknowns A_i, χ. The first N equations are obtained by matching the near-field behavior of the outer solution as $\mathbf{x} \to \mathbf{x}_j$ with the far-field behavior of the corresponding inner solution (7.2.4). After cancellation of the logarithmic terms, we have

$$-\pi A_j R_j - \pi \sum_{i \neq j} A_i G_{ji} + \chi = \frac{A_j}{\mu_j}, \tag{7.2.12}$$

for $j = 1, \ldots, N$, where $G_{ji} \equiv G(\mathbf{x}_j, \mathbf{x}_i)$ and $R_j \equiv R(\mathbf{x}_j, \mathbf{x}_j)$. The remaining equation is obtained by noting that $\nabla^2 \tau(\mathbf{x}) = -\pi \sum_{j=1}^{N} A_j \nabla^2 G(\mathbf{x}, \mathbf{x}_j)$ and hence

$$\pi |\Omega|^{-1} \sum_{j=1}^{N} A_j = \frac{1}{D}. \tag{7.2.13}$$

In the case of a single absorbing window of arc length 2ε ($d = 1/2$), Eqs. (7.2.12) and (7.2.13) are easily solved to give $A_1 = \Omega|/\pi D$, so that we obtain the equations

$$\tau(\mathbf{x}) \sim \frac{|\Omega|}{D}\left[-\frac{1}{\pi}\log(\varepsilon/2) + R(\mathbf{x}_1,\mathbf{x}_1) - G(\mathbf{x},\mathbf{x}_1)\right], \tag{7.2.14}$$

$$\overline{\tau} \sim \frac{|\Omega|}{D}\left[-\frac{1}{\pi}\log(\varepsilon/2) + R(\mathbf{x}_1,\mathbf{x}_1)\right]. \tag{7.2.15}$$

All that remains is to calculate the regular part of the Neumann Green's function $R(\mathbf{x},\mathbf{x}_j)$, which will depend on the geometry of the domain Ω. In certain cases such as the unit disk or a rectangular domain, explicit formulae for R can be obtained; otherwise numerical methods are required [267, 511, 594, 595]. The Green's function for a unit disk when the source $\mathbf{x}_j$ is on the unit circle has the well-known formula

$$G(\mathbf{x},\mathbf{x}_j) = -\frac{1}{\pi}\log|\mathbf{x}-\mathbf{x}_j| + \frac{|\mathbf{x}|^2}{4\pi} - \frac{1}{8\pi}.$$

It immediately follows that $R(\mathbf{x}_1,\mathbf{x}_1) = 1/8\pi$ (since $|\mathbf{x}_1|^2 = 1$) and

$$\overline{\tau} = \frac{1}{D}[-\log(\varepsilon) + \log 2 + 1/8].$$

For a rectangular domain of width L_2 and height L_1, the Green's function can be solved using separation of variables and expanding the result in terms of logarithms (see [70, 511]).

7.2.2 Diffusion to a Small Target

In the derivation of diffusion-limited reaction rates (Sect. 2.4), it was assumed that diffusion of the background reactants occurs in an unbounded domain with a uniform concentration at infinity. The analysis becomes considerably more involved when reactions occur in a bounded domain. Recently, Straube et al. [613] have shown how methods similar to the analysis of the narrow escape problem can be used to determine the reaction rate in the asymptotic limit that the target is much smaller than the domain size. Here we sketch the basic steps of their analysis. Consider a target disk Ω_ε of radius $\varepsilon \ll 1$ and center $\mathbf{x}_0$ that is located in the interior of a rectangular domain Ω of size $\mathcal{O}(1)$ (see Fig. 7.11). The calculation of the reaction rate can be formulated in terms of the solution to the following diffusion equation:

$$\frac{\partial c(\mathbf{x},t)}{\partial t} = D\nabla^2 c(\mathbf{x},t), \quad \mathbf{x} \in \Omega\backslash\Omega_\varepsilon \tag{7.2.16}$$

with $\partial_n c = 0$ on the exterior boundary $\partial\Omega$ and $c = 0$ on the interior boundary $\partial\Omega_\varepsilon$. The initial condition is taken to be $c(\mathbf{x},0) = 1$. Following Straube et al. [613], we seek a solution in the form of an eigenfunction expansion,

$$c(\mathbf{x},t) = \sum_{j=0}^{\infty} c_j \phi_j(\mathbf{x}) \mathrm{e}^{-\lambda_j D t} \tag{7.2.17}$$

where the eigenfunctions $\phi_j(\mathbf{x})$ satisfy the Helmholtz equation

$$\nabla^2 \phi_j + \lambda_j \phi_j = 0, \quad \mathbf{x} \in \Omega \backslash \Omega_\varepsilon \tag{7.2.18}$$

subject to the same boundary conditions as $c(\mathbf{r},t)$. The eigenfunctions are orthogonalized as

$$\int_{\Omega\backslash\Omega_\varepsilon} \phi_i(\mathbf{x})\phi_j(\mathbf{x}) d\mathbf{x} = \delta_{i,j}. \tag{7.2.19}$$

The initial condition then implies that

$$c_j = \int_{\Omega\backslash\Omega_\varepsilon} \phi_j(\mathbf{x}) d\mathbf{x}. \tag{7.2.20}$$

Taking the limit $\varepsilon \to 0$ results in an eigenvalue problem in a rectangular domain without a hole. It is well known that the eigenvalues are ordered as $\lambda_0 = 0 < \lambda_1 \leq \lambda_2 \leq \ldots$. This ordering will persist when $0 < \varepsilon \ll 1$ so that in the long-time limit, the solution will be dominated by the eigenmode with the smallest eigenvalue:

$$c(\mathbf{x},t) \sim c_0 \phi_0(\mathbf{x}) \mathrm{e}^{-\lambda_0 D t}. \tag{7.2.21}$$

The time-dependent flux is then

$$J(t) = D c_0 \mathrm{e}^{-\lambda_0 D t} \int_0^{2\pi} \left(r \frac{\partial \phi_0}{\partial r} \right) \Big|_{r=\varepsilon} d\theta. \tag{7.2.22}$$

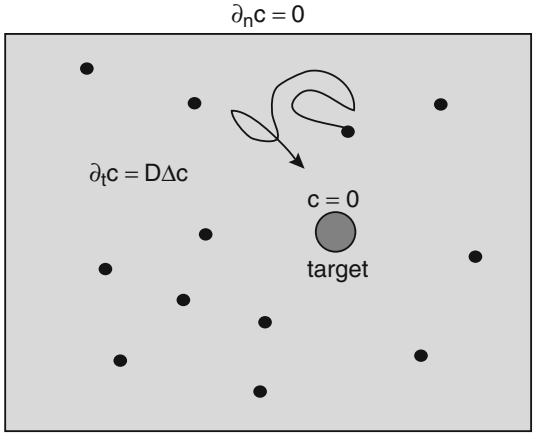


Fig. 7.11: Diffusion-limited reaction rate for a small target in a bounded domain

For small ε, the principal eigenvalue λ_0 of the Helmholtz operator has an infinite logarithmic expansion [613, 671]:

$$\lambda_0 = \nu\Lambda_1 + \nu^2\Lambda_2 + \ldots, \quad \nu = -\frac{1}{\log\varepsilon}. \tag{7.2.23}$$

Moreover, the eigenfunction $\phi_0(\mathbf{x})$ develops a boundary layer in a neighborhood of the target, where it changes rapidly from zero on the boundary $\partial\Omega_\varepsilon$ to a value of $\mathcal{O}(1)$ away from the target. This suggests dividing the domain into inner and outer regions and using matched asymptotics along analogous lines to the study of the narrow escape problem. The logarithmic expansion of λ_0 implies that the right-hand side of the rescaled eigenvalue equation is of $\mathcal{O}(\varepsilon^2\nu^2) = o(\nu^k)$ for all $k \geq 0$. Thus, to logarithmic accuracy, it follows that the inner problem with stretched coordinates $\mathbf{y} = \mathbf{x}/\varepsilon$ is

$$\nabla^2\varphi(\mathbf{y}) = 0, \quad \mathbf{y} \in \mathbb{R}^2\backslash S^1,$$

where S^1 is the unit circle centered about the origin and $\varphi = 0$ on $|\mathbf{y}| = 1$. Hence, $\varphi(\mathbf{y}) = A\log|\mathbf{y}|$ and the inner solution has the far-field behavior

$$\varphi \sim A\log(|\mathbf{x}-\mathbf{x}_0|/\varepsilon). \tag{7.2.24}$$

The outer solution satisfies the equation

$$\begin{aligned}
&\nabla^2\phi_0 + \lambda_0\phi_0 = 0, \quad \mathbf{x} \in \Omega\backslash\{\mathbf{x}_0\},\\
&\phi_0 \sim A\log(|\mathbf{x}-\mathbf{x}_0|/\varepsilon), \quad \mathbf{x} \to \mathbf{x}_0,\\
&\int_\Omega \phi_0^2(\mathbf{x})d\mathbf{x} = 1.
\end{aligned}$$

The outer problem can be solved in terms of the Neumann Green's function for the Helmholtz equation:

$$\nabla^2 G(\mathbf{x},\mathbf{x}_0;\lambda_0) + \lambda_0 G(\mathbf{x},\mathbf{x}_0;\lambda_0) = -\delta(\mathbf{x}-\mathbf{x}_0), \quad x \in \Omega \tag{7.2.25a}$$

$$\partial_n G(\mathbf{x},\mathbf{x}_0;\lambda_0) = 0, \quad \mathbf{x} \in \partial\Omega \tag{7.2.25b}$$

$$G(\mathbf{x},\mathbf{x}_0;\lambda_0) \sim -\frac{1}{2\pi}\log|\mathbf{x}-\mathbf{x}_0| + R(\mathbf{x}_0,\mathbf{x}_0;\lambda_0), \ \mathbf{x} \to \mathbf{x}_0. \tag{7.2.25c}$$

That is,

$$\phi_0(\mathbf{x}) = -2\pi A G(\mathbf{x},\mathbf{x}_0;\lambda_0). \tag{7.2.26}$$

Matching the near-field behavior of the outer solution with the far-field behavior of the inner solution then yields a transcendental equation for the principal eigenvalue:

$$R(\mathbf{x}_0,\mathbf{x}_0;\lambda_0) = -\frac{1}{2\pi\nu}. \tag{7.2.27}$$

Finally, the normalization condition for ϕ_0 determines the amplitude A according to

$$4\pi^2 A^2 \int_\Omega G(\mathbf{x},\mathbf{x}_0;\lambda_0)^2 d\mathbf{x} = 1. \tag{7.2.28}$$

Since $0 < \lambda_0 \ll 1$ for a small target, the Green's function has the expansion

$$G(\mathbf{x},\mathbf{x}_0;\lambda_0) = -\frac{1}{\lambda_0|\Omega|} + G_1(\mathbf{x},\mathbf{x}_0) + \lambda_0 G_2(\mathbf{x},\mathbf{x}_0) + \mathcal{O}(\lambda_0^2)$$

with $\int_\Omega G_j(\mathbf{x},\mathbf{x}_0)d\mathbf{x} = 0$. Substituting this expansion into Eq. (7.2.28) shows that to leading order in λ_0,

$$A \approx \frac{\sqrt{|\Omega|}\lambda_0}{2\pi}. \tag{7.2.29}$$

Similarly, Eqs. (7.2.20) and (7.2.26) imply that

$$c_0 = -2\pi A \int_\Omega G(\mathbf{x},\mathbf{x}_0;\lambda_0)d\mathbf{x} \approx \frac{2\pi A}{\lambda_0}. \tag{7.2.30}$$

The regular part $R(\mathbf{x},\mathbf{x}_0;\lambda_0)$ can also be expanded in terms of λ_0. Hence, neglecting terms of $\mathcal{O}(\lambda_0)$ and higher, substitute $R(\mathbf{x},\mathbf{x}_0;\lambda_0) \approx -(\lambda_0|\Omega|)^{-1} + R_1(\mathbf{x},\mathbf{x}_0)$ into Eq. (7.2.27). This yields a linear equation for λ_0 such that

$$\lambda_0 \approx \frac{2\pi\nu}{|\Omega|}\frac{1}{1+2\pi\nu R_1(\mathbf{x}_0,\mathbf{x}_0)}. \tag{7.2.31}$$

We now have all the components necessary to determine the time-dependent reaction rate. That is, substituting the inner solution $\phi_0(\mathbf{x}) = A\log(r/\varepsilon)$, $r = |\mathbf{x}-\mathbf{x}_0|$, into Eq. (7.2.22), and using Eqs. (7.2.29) and (7.2.30), yields the result

$$J(t) \approx D|\Omega|\lambda_0 \mathrm{e}^{-\lambda_0 Dt}, \quad \lambda_0 = \frac{2\pi\nu}{\Omega} + O(\nu^2). \tag{7.2.32}$$

7.3 Membrane Transport Through Nanopores and Channels

As we mentioned in the introduction to this chapter, an important cellular transport process is the passage of molecules through membrane channels or pores. Here we will focus on passive aspects of membrane transport, which are characterized by diffusion in confined domains. The details of the process will depend on the relative size of the molecules compared to the channel and any interactions between the diffusing particles and the channel itself. A major consequence of confinement is that it restricts the number of degrees of freedom available to a molecule, which leads to strong entropic effects. We will explore the effects of confinement in three distinct cases.

(i) *Diffusion of ions or lipids through a narrow channel.* Here changes in the motion of a molecule occur mainly in the axial direction along the channel, whereas local equilibrium is rapidly reached in the transverse directions. Thus transport is quasi-one-dimensional and the effects of the boundaries of the channel can be incorporated by introducing an effective (entropic) energy barrier into the dynamics of a Brownian particle, leading to the so-called Fick–Jacobs equation [82, 83, 297, 316, 547, 708]. Typically a 3D narrow channel is represented by a cylinder that extends axially in the x-direction and has a periodically varying cross section that is rotationally symmetric about the x-axis (see Fig. 7.12a). Denoting the space-dependent radius of period L by $w(x)$, with $w(x+L) = w(x)$ for all x, the cross-section varies as $A(x) = \pi w(x)^2$. In the case of a corresponding 2D channel, $w(x)$ represents the half-width of the channel.

(ii) *Single-file diffusion.* An extreme version of confined diffusion along a channel is single-file diffusion, in which the channel is so narrow that particles cannot pass each other. In other words, the longitudinal motion of each particle is hindered by the presence of its neighbors, which act as moving obstacles (see Fig. 7.12b). Hence, interparticle interactions can suppress Brownian motion and lead to subdiffusive behavior [22, 386, 504, 621].

(iii) *Translocation of polymers through a pore.* Polymer translocation through a membrane pore plays an important role in a number of cellular processes, including the transport of RNA across nuclear pores, virus infection of cells, and DNA packaging into viral capsids [3]. There are also an increasing number of technological applications, ranging from drug delivery to biochemical sensors, which have been driven by significant progress in experimental studies of translocation at the single-molecule level [463]. Translocation of biopolymers in vivo is typically facilitated by interactions with the channel or specialized proteins (chaperones), whereas translocation in vitro is driven by the application of external electrical fields. Since the resulting voltage difference across the pore would normally cause the flow of ions, one can measure when

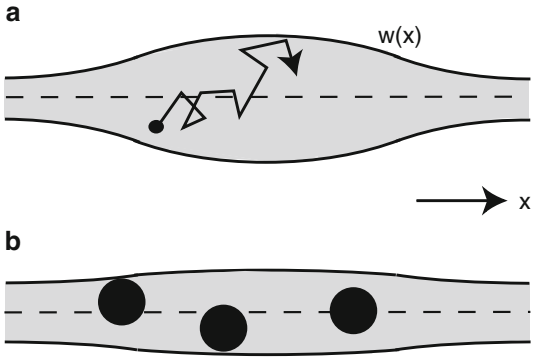


Fig. 7.12: Confined diffusion in a narrow cylindrical channel with a periodically modulated boundary $w(x)$ in the axial direction. (**a**) Small diffusing particle. (**b**) Single-file diffusion

a polymer enters the pore since it partially blocks the path of the ions, resulting in a significant decrease in the ionic current. Thus the frequency and duration of translocation events can be recorded. From a theoretical perspective the translocation process also involves diffusion past an entropic barrier, in this case arising from the fact that a free polymer has many more configurational states than one that is threaded through a pore [117, 462, 493, 615].

7.3.1 Confined Diffusion and the Fick–Jacobs Equation

We begin by deriving the Fick–Jacobs equation for a Brownian particle diffusing in a 2D channel as shown in Fig. 7.12a. We follow the particular derivation of [708]. It is assumed that the channel walls at $y = \pm w(x)$ confine the motion of the particle but do not exchange energy with it. Thus the probability flux normal to the boundary is zero. This condition can be imposed by introducing a confining potential $U(x,y)$ such that $U(x,y) = 0$ for $|y| < w(x)$ and $U(x,y) = \infty$ for $|y| \geq w(x)$. Let $p(x,y,t)$ denote the probability that the particle is located at position $\mathbf{x} = (x,y)$ at time t with periodic boundary conditions in the longitudinal direction, $p(x+L,y,t) = p(x,y,t)$. For a general potential $U(x,y)$, the 2D FP equation takes the form

$$\frac{\partial p}{\partial t} = -\frac{1}{\gamma}\left[\frac{\partial [F_x p]}{\partial x} + \frac{\partial [F_y p]}{\partial y}\right] + D_0\left[\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2}\right],$$

where $F_x = -\partial_x U, F_y = -\partial_y U$. Using the Einstein relations $D_0\gamma = k_B T = \beta^{-1}$, the FP equation can be rewritten as

$$\begin{aligned}\frac{\partial p}{\partial t} &= D_0 \frac{\partial}{\partial x} \mathrm{e}^{-\beta U(x,y)} \frac{\partial}{\partial x} \mathrm{e}^{\beta U(x,y)} p(x,y,t) \\ &\quad + D_0 \frac{\partial}{\partial y} \mathrm{e}^{-\beta U(x,y)} \frac{\partial}{\partial y} \mathrm{e}^{\beta U(x,y)} p(x,y,t). \end{aligned} \tag{7.3.1}$$

In order to reduce to a 1D equation, first integrate both sides of the FP equation with respect to the transverse coordinate y:

$$\frac{\partial P(x,t)}{\partial t} = D_0 \frac{\partial}{\partial x} \int_{-w(x)}^{w(x)} \mathrm{e}^{-\beta U(x,y)} \frac{\partial}{\partial x} \mathrm{e}^{\beta U(x,y)} p(x,y,t) dy,$$

where $P(x,t)$ is the reduced probability density

$$P(x,t) = \int_{-w(x)}^{w(x)} p(x,y,t) dy. \tag{7.3.2}$$

The major step in the reduction is to assume that the probability density reaches equilibrium in the transverse direction. That is, $p(x,y,t)$ is assumed to factorize as follows:

$$p(x,y,t) \approx P(x,t)\rho(x,y), \tag{7.3.3}$$

where $\rho(x,y)$ is a normalized Boltzmann–Gibbs probability density (Sect. 1.4):

$$\rho(x,y) = \frac{\mathrm{e}^{-\beta U(x,y)}}{A_0 \mathrm{e}^{-\beta \mathscr{U}(x)}}, \quad \mathrm{e}^{-\beta \mathscr{U}(x)} = \frac{1}{A_0}\int_{-w(x)}^{w(x)} \mathrm{e}^{-\beta U(x,y)} dy, \tag{7.3.4}$$

where $A_0 = 2\int_0^L w(x)dx$ and $\mathscr{U}(x)$ interpreted as an effective x-dependent barrier potential or free energy. Under this factorization the averaged FP equation becomes

$$\frac{\partial P(x,t)}{\partial t} \approx D_0 \frac{\partial}{\partial x}\mathrm{e}^{-\beta \mathscr{U}(x)} \frac{\partial}{\partial x}\mathrm{e}^{\beta \mathscr{U}(x)} P(x,t). \tag{7.3.5}$$

This holds for a general potential energy function $U(x,y)$ [526]. If U is now taken to be the confining potential of the channel boundary, then $\mathrm{e}^{-\beta \mathscr{U}(x)} = 2w(x)/A_0 \equiv \sigma(x)$ and we obtain the Fick–Jacobs equation

$$\frac{\partial P(x,t)}{\partial t} = D_0 \frac{\partial}{\partial x}\sigma(x)\frac{\partial}{\partial x}\frac{P(x,t)}{\sigma(x)}. \tag{7.3.6}$$

The same equation is obtained in 3D with $\sigma(x) = A(x)/A_0$ with $A(x) = \pi w(x)^2$ and $A_0 = \pi\int_0^L w(x)^2 dx$. In the physics literature $\mathscr{U}(x)$ is usually referred to as an *entropic barrier*, since confinement reduces the volume of the phase space of available states accessible to the particle.[1] The Fick–Jacobs equation is valid provided that $|w'(x)| \ll 1$. However, it has been shown that the introduction of an x-dependent diffusion coefficient into the Fick–Jacobs equation can considerably increase the accuracy of the reduced FP equation and thus extend the domain of validity [316, 526, 708]:

$$D(x) = \frac{D_0}{[1+w'(x)^2]^\alpha}, \tag{7.3.7}$$

with $\alpha = 1/3, 1/2$ for 2D and 3D, respectively.

As it stands, the Fick–Jacobs equation (7.3.6) represents a particle diffusing in a 1D periodic potential $\mathscr{U}(x)$, which means that the mean velocity of the particle is zero. On the other hand, net transport of the particle through the channel does occur in the presence of a constant external force F_0 in the x-direction. Equation (7.3.6) still holds, except that now $\mathscr{U}(x) = -F_0 x - k_B T \ln\sigma(x)$, which yields the classical

[1] The effective free energy $\mathscr{U}(x) = -k_B T \log[A(x)/A_0]$ reflects the existence of an entropic barrier to diffusion [526]. That is, using the standard definition of free energy $\mathscr{U} = E_0 - TS$, where E_0 is internal energy and S is the entropy, it follows that $S(x) \sim \log A(x)$ where $A(x)$ is the cross-sectional area of the channel at x. This is consistent with the definition of entropy in terms of the logarithm of the number of microstates. That is, in equilibrium there is a uniform probability density ρ_0 in the channel, so that the equilibrium x-dependent density $P_{\mathrm{eq}}(x) = \rho_0 A(x)/A_0$ and the number of microstates available to a diffusing particle at location x is proportional to the area of the channel.

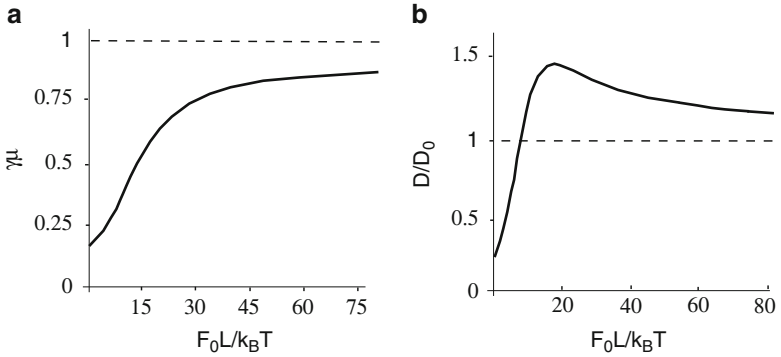


Fig. 7.13: Illustrative sketches of how mobility and diffusivity vary with nondimensionalized applied force F_0L/k_BT in the case of a 2D channel with a sinusoidally varying half-width Eq. (7.3.10). (**a**) Effective mobility μ in units of γ. In the limit $F_0 \to \infty$, $\mu \to \gamma^{-1}$. (**b**) Diffusion coefficient D in units of free diffusivity D_0. In the limit $F_0 \to \infty$, $D \to D_0$. Sketches are based on numerical results of [527] for $a = L/2\pi$ and $b = 1.02$

problem of Brownian motion in a periodic potential with tilt [82, 253, 530, 612], which was analyzed in Sect. 4.2. Given the mean and variance of the particle position in the long-time limit, one can define the drift mobility and diffusion coefficient of the particle according to

$$\mu(F_0) \equiv \frac{\langle \dot{X} \rangle}{F_0}, \quad \langle \dot{X} \rangle = \lim_{t \to \infty} \frac{\langle X(t) \rangle}{t}, \tag{7.3.8}$$

and

$$D(F_0) = \lim_{t \to \infty} \frac{\langle X(t)^2 \rangle - \langle X(t) \rangle^2}{2t}. \tag{7.3.9}$$

Note that the relationship between $\langle \dot{X} \rangle$ and the long-time limit of $\langle X(t) \rangle / t$ is a consequence of ergodicity [529] (see Sect. 4.2). The force dependence of the mobility and diffusion coefficient have been studied both analytically and numerically in the case of a sinusoidal boundary function [82, 527]

$$w(x) = a[\sin(2\pi x/L) + b], \quad a > 0, b > 1. \tag{7.3.10}$$

The basic results are sketched in Fig. 7.13. A number of interesting observations emerge from this study. First, the mobility only depends on the temperature via the dimensionless parameter F_0L/k_BT. Hence, increasing the temperature reduces the mobility. Second, as the force is increased the effective diffusion coefficient $D(F_0)$ exceeds the free diffusion coefficient D_0. Finally, note that for certain forms of $\sigma(x)$, one can find exact solutions of the Fick–Jacobs equation by a change of variables [543] (see Ex. 7.3).

The Fick–Jacobs equation represents diffusion through a narrow channel in terms of a 1D overdamped Brownian particle moving in an effective potential $\mathscr{U}(x)$ that arises from confinement. Such a 1D model has also been the starting point for a series of studies of channel-facilitated membrane transport, where now $\mathscr{U}(x)$ reflects the constructive role of attractive interactions between permeating particles and proteins forming the channel pore. In a series of studies [33–35], mixed boundary conditions are assumed at the ends $x = 0, L$ of the channel: $J(0,t) = -\kappa_0 P(0,t)$ and $J(L,t) = -\kappa_L P(L,t)$. The probability of crossing the channel and the mean time in the channel is then calculated in terms of FPTs and splitting probabilities. It can be shown that there is an optimal form of the interaction potential that maximizes the flux through the channel and involves a play-off between increasing the translocation probability through the channel and decreasing the average time particles spend in the channel [33] (see also the entropic gate model in Sect. 7.3.2). For a complementary approach to studying channel-facilitated transport that is based on spatially discrete stochastic site-binding models, see [112, 345].

Finally, note that a variety of models have been developed to analyze ion permeation through narrow channels. Broadly speaking, these models can be divided into three classes [261, 387, 546]: (i) Brownian dynamics models based on a Langevin description of the full 3D dynamics of ion motion through channels, which take into account both ion–ion interactions and ion channel interactions. (ii) Continuum mean-field models based on the *Poisson–Nernst–Planck* equation. The latter treats a channel as a continuous medium, and the ionic current is determined by coupling the Nernst–Planck electrodiffusion equation for the flux of charged particles in the presence of a concentration gradient and an electric field (see Ex. 2.4) with the Poisson equation describing how the distribution of charges generates an effective mean-field potential [464, 575]. (iii) Barrier models in which ions are localized to specific regions of the channel via local potentials and the kinetics are represented by the rate constants for hopping between these regions and between the channel and the bulk [261]. A simple barrier model of ion permeation is considered in Ex. 7.4. (In Chap. 3 we considered the stochastic gating of ion channels and their role in membrane excitability, assuming a simple linear (ohmic) relationship between membrane voltage and current flow through an open ion channel.)

7.3.2 Nuclear Transport

The nucleus of eukaryotes is surrounded by a protective nuclear envelope (NE) within which are embedded NPCs (see Fig. 7.14). The NPCs are the sole mediators of exchange between the nucleus and cytoplasm. In general small molecules of diameter ~ 5 nm can diffuse through the NPCs unhindered, whereas larger molecules up to around 40 nm in diameter are excluded unless they are bound to a family of soluble protein receptors known as karyopherins (kaps) (see the reviews [403, 545, 641]). Within the cytoplasm kap receptors bind cargo to be imported via a nuclear localization signal (NLS) that results in the formation of a kap–cargo

complex. This complex can then pass through an NPC to enter the nucleus. A small enzyme RanGTP then binds to the kap, causing a conformational change that releases the cargo. The sequence of events underlying the import of cargo is shown

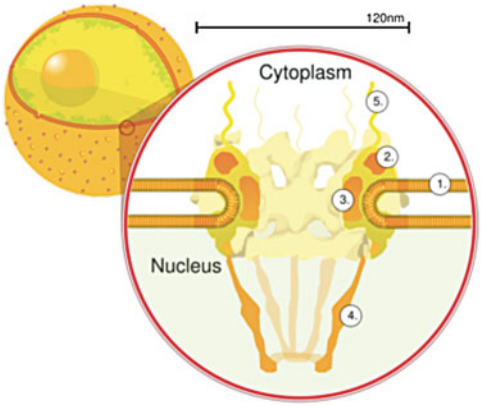


Fig. 7.14: The nuclear pore complex. *1* Nuclear envelope. *2* Outer ring. *3* Spokes. *4* Basket. *5* Filaments. Each of the eight protein subunits surrounding the actual pore (the outer ring) projects a spoke-shaped protein into the pore channel. (Public domain figure from Wikimedia.)

in Fig. 7.15a. In the case of cargo export from the nucleus, kaps bind to cargo with a nuclear export signal (NES) in the presence of RanGTP, and the resulting complex passes through the NPC. Once in the cytoplasm, RanGTP undergoes hydrolysis to form RanGDP, resulting in the release of the cargo. The export process is illustrated in Fig. 7.15b. Finally, RanGDP is recycled to the nucleus by another molecule NFT2 and is reloaded with GTP to begin another import/export cycle. This cycle allows a single NPC to support a very high rate of transport on the order of 1,000 translocations/seconds [533]. Since the transportation cycle is directional and accumulates cargo against a concentration gradient, an energy source combined with a directional cue is required. Both of these are provided by the hydrolysis of RanGTP and the maintenance of a concentration gradient of RanGTP across the NPC. The RanGTP gradient is continuously regenerated by GTP hydrolysis in the cytoplasm, translocation of RanGTD into the nucleus by NFT2, and replacement of GDP by GTP in the nucleus. It is important to note that the energy generated from RanGTP hydrolysis is ultimately used to create a concentration gradient of RanGTP between the nucleus and cytoplasm, so that the actual translocation across the NPC occurs purely via diffusion.

Although the above basic picture is now reasonably well accepted, the detailed mechanism underlying facilitated diffusion of kap–cargo complexes within the NPC is still not understood. The NPC is composed of about 30 distinct proteins known collectively as nucleoporins (nups). It has emerged in recent years that individual nups are directly related to a number of human diseases including influenza and cancers such as leukemia [130], as well as playing an important role in viral infections

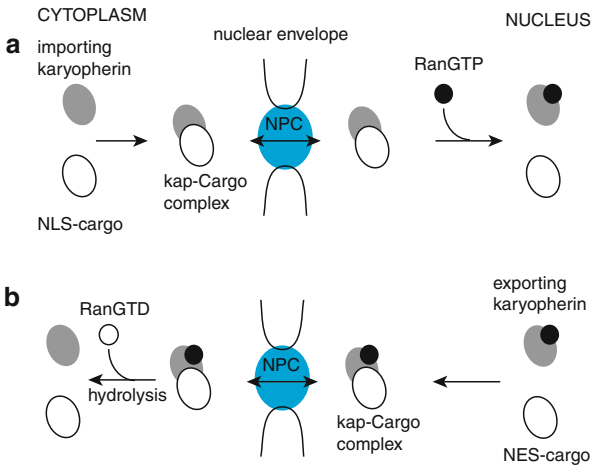


Fig. 7.15: Schematic illustration of the (**a**) import and (**b**) export process underlying the karyopherin-mediated transportation of cargo between the nucleus and cytoplasm via a nuclear pore complex (NPC). See text for details

by providing docking sites for viral capsids [679]. Associated with many of the nups are natively unfolded phenylalanine–glycine (FG) repeats, known collectively as FG-repeats. The FG-repeats set up a barrier to diffusion for large molecules so that the key ingredient in facilitated diffusion through the NPC is the interaction between kap receptors with the FG-repeats. In essence, the major difference between the various theoretical models of NPC transport concerns the built-in assumptions regarding the properties and spatial arrangements of FG-repeats within the NPC and the nature of interactions with kaps during translocation through the NPC [27].

Entropic Gate Model

Recall from Sect. 7.3.1 that a macromolecule diffusing in a confined geometry (such as a nuclear pore) experiences an entropic barrier due to excluded volume effects. Within the NPC this would be enhanced by the densely packed FG-repeats. One way to counteract the effects of the entropic barrier is for the kaps to have an affinity for and bind to the FG-repeat regions [545, 704], thus lowering the effective free energy of the cargo complex within the NPC. The degree of affinity has to be sufficiently high to overcome the entropic barrier but not too high; otherwise the complex can be trapped within the NPC and the rate of translocation would be too small. One possible solution is to have a large number of low-affinity binding sites within the nuclear pore. Recently, a mathematical model for the effects of binding on diffusion within the NPC has been developed by [704], based on diffusion through an effective energy landscape, which approximates the effects of multiple binding sites when the binding/unbinding rates are relatively fast compared to the diffusion rate (see

below). The simplest version of the model is illustrated in Fig. 7.16 for the case of nuclear import. The effective potential energy $U(x)$ is taken to be a flat potential well of depth E along an NPC and zero outside the NPC. Absorbing boundary conditions are placed at the points $x = 0, L$ a distance R from either side of the NPC, which has length $L - 2R$. The absorbing boundary conditions represent a cargo-complex

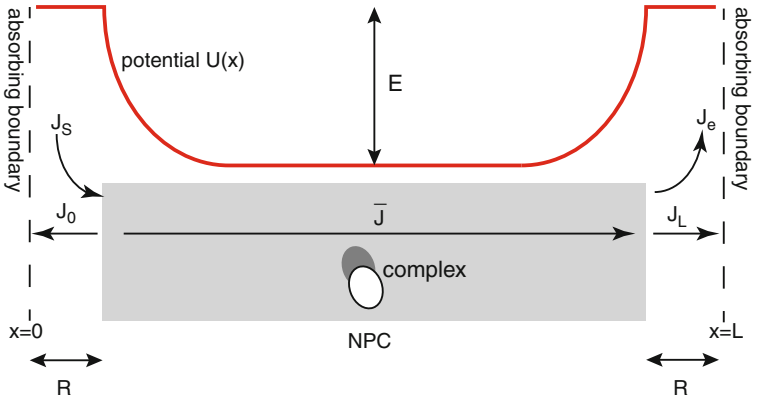


Fig. 7.16: Sketch of model of [704]. Transport of cargo complex through the NPC is modeled as diffusion in an energy landscape. See text for details

returning to the cytoplasm at $x = 0$ or entering the nucleus at $x = L$, and diffusing away. Diffusion within the NPC is described by a standard Smoluchowski equation for the density of cargo complexes $\rho(x)$, $x = [0, L]$:

$$\frac{\partial \rho}{\partial t} = -\frac{\partial J}{\partial x}, \quad J = -D\frac{\partial \rho}{\partial x} - D\rho\frac{\partial U}{\partial x}, \tag{7.3.11}$$

with U measured in units of $k_B T$. This equation is supplemented by the absorbing boundary conditions $\rho(0) = \rho(L) = 0$.

The steady-state solution is obtained by assuming that there are constant fluxes J_0 in $[0, R]$, J_L in $[L-R, L]$, and $\overline{J}$ in $[R, L-R]$ with $J_0 < 0$. These fluxes are related according to $J_S = \overline{J} - |J_0|$ and $\overline{J} = J_L + J_e$, where J_S is the total flux of complexes injected into the NPC from the cytoplasm, which is proportional to the density of complexes in the cytoplasm, and J_e denotes the flux due to active removal of complexes from the nucleus end of the NPC by RanGTP. The latter depends on the number of complexes at the nuclear exit, the rate J_{ran} at which RanGTP molecules hit the exit: $J_e = J_{\text{ran}}\rho(L-R)R$. The steady-state rate of transport $\overline{J}$ can now be determined by solving for $\rho(x)$ in terms of $J_0, J_L, \overline{J}$ in each of the three domains and imposing continuity of the density at $x = R$ and $x = R - L$ (see Ex. 7.5). The result is that the fraction of complexes reaching the nucleus is given by [704]

$$P = \frac{\overline{J}}{J_S} = \left[1 + \frac{1}{1+K} + \frac{1}{R}\int_R^{L-R} \mathrm{e}^{U(x)} dx\right]^{-1}, \tag{7.3.12}$$

with $K = J_{\text{ran}}R^2/D$. It follows that for a sufficiently deep well (large E), where the integral term is negligible, and for sufficiently large K (large J_{ran}), the probability of translocation is $P \approx 1$. On the other hand, if K is small so that RanGTP does not facilitate entry of complexes into the nucleus, then $P_{\max} = 0.5$. As previously indicated, it is not possible to arbitrarily increase the affinity of binding sites

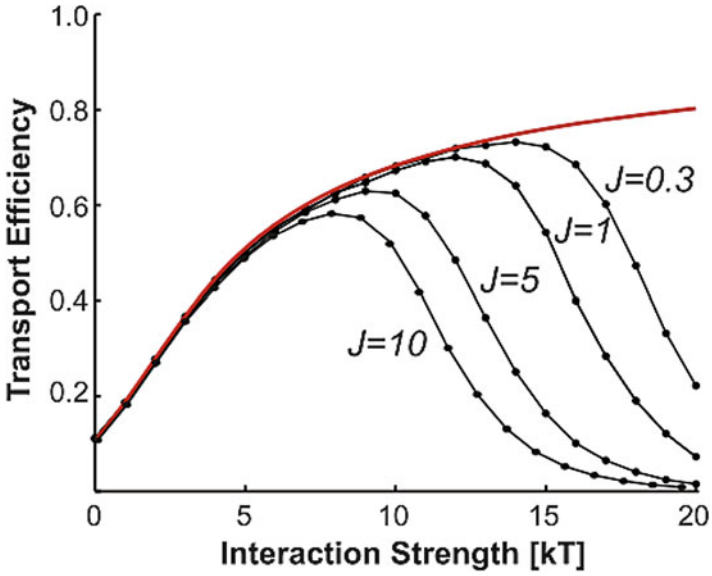


Fig. 7.17: Results of Monte Carlo simulations carried out by Zilman et al. [704] on their entropic gate model. Transport efficiency (probability P to reach the nucleus) is plotted as a function of the NPC interaction strength E. The unimodal curves correspond to four different values of the entrance rate J (in units of $10^{-4}16D/R^2$) and have a peak at a specific value of E, which provides a mechanism of selectivity. The envelope is the theoretical curve calculated from Eq. (7.3.12). RanGTP activity in the nucleus is fixed by setting $J_{\text{ran}}L^2/N^2D = 1.5$ (Adapted from Zilman et al. [704])

and thus the well depth E, since this will lead to trapping of the complexes so that they accumulate within the NPC, resulting in molecular crowding and an unrealistically long time for an individual molecule to pass through the NPC. Thus there is some optimal well depth that balances an increase of transport probability P with increased time spent in the NPC [704]. Finally, note that the model is robust with regard to the particular shape of the potential well. For example, one could represent transport through the NPC as diffusion in an array of overlapping potential wells that represent flexible FG-repeat regions. The shape of each well will depend on the number and affinity of binding sites on each FG-repeat, and the degree of flexibility of the polymers which will determine the entropic costs of bending and stretching the FG-repeats. Results of Monte Carlo simulations are shown in Fig. 7.17. One difference from the analytical model is that only one complex is allowed to occupy a

given site in the discretized model, which takes into account molecular crowding. However, this effect is small for small interaction strength E, since the density of complexes in the NPC is low, and there is good agreement between numerics and theory.

It is straightforward to show that for relatively fast binding/unbinding, the multi-well potential can be replaced by a single well along the lines of Fig. 7.16. Suppose that inside the NPC there are N binding sites labeled $i = 1, \ldots, N$ with corresponding potentials $U_i(x)$. Let $U_0(x)$ be the potential of an unbound complex. Denoting the probability density of complexes bound to the jth site by $\rho_i(t)$ and the density of unbound complexes by $\rho_0(t)$, we have the system of equations

$$\frac{\partial \rho_0}{\partial t} = D_0 \frac{\partial}{\partial x} \mathrm{e}^{-U_0(x)} \frac{\partial}{\partial x} \mathrm{e}^{U_0(x)} \rho_0 + \sum_{i>0} [\omega_{0i}(x)\rho_i(x) - \omega_{i0}(x)\rho_0(x)]$$
$$\frac{\partial \rho_i}{\partial t} = D_0 \frac{\partial}{\partial x} \mathrm{e}^{-U_i(x)} \frac{\partial}{\partial x} \mathrm{e}^{U_i(x)} \rho_i + [\omega_{i0}(x)\rho_0(x) - \omega_{0i}(x)\rho_i(x)].$$

Here a complex can unbind from the ith site at a rate $\omega_{0i}(x)$ and rebind to the same site or bind to a different site at a rate $\omega_{j0}(x)$. From detailed balance (see Sect. 1.4), the transition rates are related according to

$$\frac{\omega_{i0}(x)}{\omega_{0i}(x)} = \mathrm{e}^{-U_i(x)+U_0(x)}.$$

Under the assumption of fast transition rates, the densities in the different states will be at local thermodynamic equilibrium with respect to the internal states. We thus have the Boltzmann–Gibbs distribution

$$\rho_i(x) = \frac{\mathrm{e}^{-U_i(x)}}{\sum_{i=0}^{M} \mathrm{e}^{-U_i(x)}} \rho(x), \quad \rho(x) = \sum_{i=0}^{N} \rho_i(x).$$

Hence adding together the $N+1$ differential equation gives

$$\frac{\partial \rho}{\partial t} = D_0 \frac{\partial}{\partial x} \mathrm{e}^{-U(x)} \frac{\partial}{\partial x} \mathrm{e}^{U(x)} \rho,$$

where

$$U(x) = -\ln \left(\mathrm{e}^{-U_0(x)} + \sum_{i>0} \mathrm{e}^{-U_i(x)} \right).$$

Selective Phase Model

The basic assumption of these models is that the NPC can be treated as a channel filled with a hydrophobic medium consisting of a concentrated polymer solution (see Fig. 7.18); the latter is composed of the natively unfolded, flexible protein domains of FG-repeats [49, 362, 533, 534]. The FG-repeats form weak bonds with

each other suggesting that they act approximately like a weak *reversible gel*. (A gel is a jelly-like material that is mostly liquid by weight, yet behaves like a solid due to a three-dimensional cross-linked polymer network within the liquid. It is the cross-linking within the fluid that gives a gel its solid-like properties such as hardness. A gel is said to be reversible if the cross-linking is reversible.) Particles smaller than the mesh size of the network can diffuse freely through the NPC, whereas nonselective macromolecules larger than the mesh size cannot. On the other hand, kap–cargo

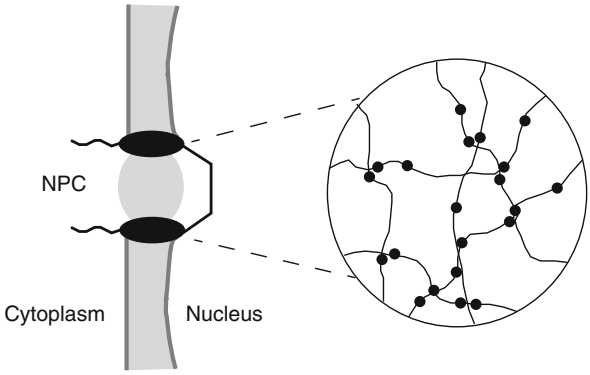


Fig. 7.18: Selective phase model [49, 533, 534], in which the FG-repeats within an NPC are treated as a reversible polymer gel. See text for details

complexes can "dissolve" in the gel due to the presence of hydrophobic domains on the surface of the kap receptors and then diffuse through the pore by breaking the weak bonds of the reversible gel [533, 534].

7.3.3 Single-File Diffusion

When a pore or channel becomes sufficiently narrow, particles are no longer able to pass each other, which imposes strong constraints on the diffusive motion. An idealized model of single-file diffusion considers a 1D collection of diffusing particles with hard-core repulsion. The many-body problem of single-file diffusion was originally tackled by relating the dynamics of the interacting system with the effective motion of a free particle [375, 386, 504]. In particular, in the case of an infinite system and a uniform initial particle density, it was shown that a tagged particle exhibits anomalous subdiffusion on long time scales, $\langle X^2(t)\rangle \sim t^{1/2}$. (On the other hand, the center of mass of the system of particles exhibits normal diffusion.) More recently, a variety of complementary approaches to analyzing single-file diffusion have been developed [22, 100, 542, 620]. Here we review the particular formulation of Barkai and Silbey [22], which develops the analysis of a tagged particle in terms of classical reflection and transmission coefficients.

Suppose that the tagged particle is initially at the origin with N particles to its left and N particles to its right (see Fig. 7.19a). The motion of each particle in the absence of hard-core interactions is taken to be overdamped Brownian motion as described by the Langevin equation (2.2.1) or the corresponding FP equation (2.2.5). As a further simplification, the potential energy function $V(x) = \int^x F(x')dx'$ is taken to be symmetric, $V(x) = V(-x)$, as is the initial distribution of particles. That is, if the initial position x_0 of a particle is drawn from $f_R(x_0)$ for $x_0 > 0$ and from $f_L(x_0)$ for $x_0 < 0$, then $f_R(x_0) = f_L(-x_0) \equiv f(x_0)$. This reflection symmetry ensures that $\langle X(t)\rangle = 0$, where $X(t)$ is the stochastic position of the tagged particle at time t. The main underlying idea is to map the many-body problem to a non-interacting one by allowing particles to pass through each other and keeping track of the particle label (see Fig. 7.19b). That is, assuming that collisions are elastic and neglecting n-body interactions for $n > 2$, it follows that when two particles collide they exchange momenta and this is represented as an exchange of particle labels. The probability density for the tagged particle to be at $X(t) = X_T$ at time t then reduces to the problem of finding the probability that the number of free particle trajectories that started at $x_0 < 0$ and are now to the right of X_T is balanced by the number of free particle trajectories that started at $x_0 > 0$ and are now to the left of X_T.

Thus, let $P_{LL}(x_0^{-j})$ ($P_{LR}(x_0^{-j})$) denote the probability that the jth free particle trajectory starting from $x_0^{-j} < 0$ at $t = 0$ is to the left (right) of X_T at time t. Similarly, let $P_{RR}(x_0^j)$ ($P_{RL}(x_0^j)$) denote the probability that the jth free particle trajectory starting from $x_0^j > 0$ at $t = 0$ is to the right (left) of X_T at time t. Let α be the net number of free particle trajectories that are on the opposite side of X_T at time t compared to their starting point (with left to right taken as positive). The associated probability distribution for α given $2N$ untagged particles is [22]

$$P_N(\alpha) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\prod_{j=1}^{N}\Gamma(\phi, x_0^{-j}, x_0^j)\mathrm{e}^{i\alpha\phi}d\phi, \tag{7.3.13}$$

where

$$\begin{aligned}\Gamma(\phi, x_0^{-j}, x_0^j) = {} & \mathrm{e}^{i\phi}P_{LR}(x_0^{-j})P_{RR}(x_0^j) + P_{LL}(x_0^{-j})P_{RR}(x_0^j) \\ & + P_{LR}(x_0^{-j})P_{RL}(x_0^j) + \mathrm{e}^{-i\phi}P_{LL}(x_0^{-j})P_{RL}(x_0^j).\end{aligned} \tag{7.3.14}$$

The integration with respect to ϕ ensures that the net number of crossings is α, that is, $\int_{-\pi}^{\pi}\mathrm{e}^{i\phi n} = \delta_{n,0}$. Since the trajectories are independent and the initial conditions are (iid) random variables, $P_N(\alpha)$ can be averaged with respect to the initial conditions to give

$$\langle P_N(\alpha)\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi}\langle\Gamma(\phi)\rangle^N\mathrm{e}^{i\alpha\phi}d\phi, \tag{7.3.15}$$

where

$$\langle\Gamma(\phi)\rangle = \left(\langle P_{RR}\rangle + \mathrm{e}^{-i\phi}\langle P_{RL}\rangle\right)\left(\langle P_{LL}\rangle + \mathrm{e}^{i\phi}\langle P_{LR}\rangle\right). \tag{7.3.16}$$

The averages $\langle P_{LR}\rangle$ can be calculated using the fundamental solution $K(x, x_0, t)$ of the corresponding FP equation (2.2.5), that is, the solution with initial condition

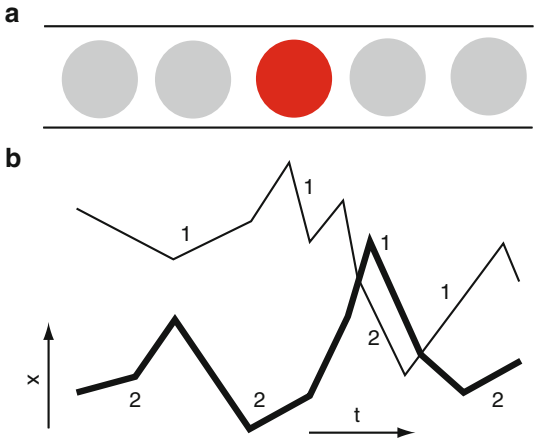


Fig. 7.19: (**a**) Single-file diffusion of a tagged particle (*darker filled circle*) surrounded by other impenetrable particles. (**b**) Equivalent noninteracting picture, in which each trajectory is treated as a noninteracting Brownian particle by keeping track of the exchange of particle label

$K(x,x_0,0)=\delta(x-x_0)$. For example,

$$\langle P_{LR}\rangle=\int_{-l}^{0}f_L(x_0)\int_{X_T}^{l}K(x,x_0,t)dxdx_0, \tag{7.3.17}$$

where $2l$ is the length of the 1D domain. For fixed $x_0<0$, $\int_{X_T}^{l}K(x,x_0,t)dx$ is the probability of being to the right of X_T at time t, which is then averaged over all initial conditions to the left of the origin.

Equation (7.3.15) takes the form of the generating function for a discrete random walk of N steps and a net displacement of α [see Eq. (2.1.4)]. Hence, for large N, application of the central limit theorem (Sect. 2.1) leads to the Gaussian approximation

$$P_N(0)\sim\frac{1}{\sqrt{2\pi N\sigma^2}}\exp(-N\mu_1^2/2\sigma^2), \tag{7.3.18}$$

where $\sigma^2=\mu_2-\mu_1^2$ and μ_1,μ_2 are the first two moments of the structure function:

$$\langle\Gamma(\phi)\rangle=1+i\mu_1\phi-\frac{1}{2}\mu_2\phi^2+\mathscr{O}(\phi^3). \tag{7.3.19}$$

Hence,

$$\mu_1=\langle P_{LR}\rangle-\langle P_{RL}\rangle,\ \sigma^2=\langle P_{RR}\rangle\langle P_{RL}\rangle+\langle P_{LL}\rangle\langle P_{LR}\rangle. \tag{7.3.20}$$

Since $\langle X(t)\rangle=0$ and N is assumed to be large, μ_1 and σ^2 can be Taylor expanded with respect to X_T about $X_T=0$. Reflection symmetry then implies that

$$\langle P_{LL}\rangle|_{X_T=0}=\langle P_{RR}\rangle|_{X_T=0}\equiv\mathscr{R},$$

$$\langle P_{LR}\rangle|_{X_T=0} = \langle P_{RL}\rangle|_{X_T=0} \equiv \mathscr{T} = 1-\mathscr{R},$$
$$\partial_{X_T}\langle P_{LR}\rangle|_{X_T=0} = -\partial_{X_T}\langle P_{RL}\rangle|_{X_T=0} \equiv \mathscr{J}.$$

The time-dependent functions $\mathscr{R}$ and $\mathscr{T}$ may be interpreted as reflection and transmission coefficients determining whether or not a free particle trajectory crosses $X_T = 0$. The resulting mean and variance are

$$\mu_1 = -2\mathscr{J}X_T + \mathscr{O}(X_T^2), \quad \sigma^2 = 2\mathscr{R}(1-\mathscr{R}) + \mathscr{O}(X_T). \tag{7.3.21}$$

Thus, $\langle P_N(\alpha)\rangle$ for $\alpha = 0$ reduces to a Gaussian distribution for the position $X(t) = X_T$:

$$P(X_T,t) = \frac{1}{\sqrt{2\pi\langle X(t)^2\rangle}}\exp\left[-\frac{X_T^2}{2\langle X(t)^2\rangle}\right], \tag{7.3.22}$$

with

$$\langle X(t)^2\rangle = \frac{\mathscr{R}(1-\mathscr{R})}{2N\mathscr{J}^2}. \tag{7.3.23}$$

Finally, using Eq. (7.3.17),

$$\mathscr{R} = \int_0^l f(x_0)\int_0^l K(x,x_0,t)dxdx_0 \tag{7.3.24}$$

$$\mathscr{J} = \int_0^l f(x_0)K(0,x_0,t)dx_0. \tag{7.3.25}$$

In the special case of zero external forces and $l\to\infty$, the fundamental solution is

$$K(x,x_0,t) = \frac{1}{\sqrt{4\pi Dt}}\mathrm{e}^{-(x-x_0)^2/4Dt}. \tag{7.3.26}$$

That is, $\lim_{t\to 0}K(x,x_0,t) = \delta(x-x_0)$ and $\partial_t K = D\partial_{xx}K$ for $t>0$, $x\in\mathbb{R}$. Taking a uniform initial distribution $f(x_0) = 1/l$ with $l\to\infty$ and fixed particle density $\rho = N/l$, one finds anomalous subdiffusion for large times t:

$$\langle X(t)^2\rangle \sim \frac{2}{\sqrt{\pi}}\frac{\sqrt{Dt}}{\rho}. \tag{7.3.27}$$

On the other hand, for particles initially centered at the origin, $f(x_0) = \delta(x_0)$, diffusion is normal

$$\langle X(t)^2\rangle \sim \frac{\pi Dt}{2N}. \tag{7.3.28}$$

In the case of a bounded domain or a Gaussian initial condition, anomalous diffusion occurs at intermediate times only [22].

7.3.4 Translocation of DNA Through a Nanopore

As our final example of membrane transport, we will consider a simple model of DNA translocation that is based on diffusion through an entropic barrier [462, 615]. There have been many subsequent improvements and extensions of the basic model, as highlighted in the recent review by Panja et al. [493]. In order to develop the theory, it will be necessary to apply the statistical mechanics of polymers introduced in Sect. 4.5. A schematic illustration of the basic physical problem is shown in Fig. 7.20. The membrane is treated as an infinitesimally thin plate separating two regions I and II with electrical potentials Φ_1 and Φ_2, respectively. The DNA polymer is represented as a freely jointed chain of N links, each of length $a \approx 100\,\text{nm}$ (the Kuhn length). For simplicity, the polymer is translocated through a pore in the membrane as a single strand. Suppose that there are $N-m$ segments in region I and m segments in region II, and in each region the chain is modeled as a random walk of m or $N-m$ segments pinned at the pore at one end. A final simplification of the model is that translocation is taken to be sufficiently slow so that at each step, the chains have enough time to reach thermodynamic equilibrium. (This assumption is relaxed in [117].) Note that the model differs considerably from the translocation ratchet analyzed in Sect. 4.2.2. The latter treats the polymer as rigid and focuses on the rectifying effects of chaperones; a discrete model of chaperone-assisted translocation is considered in Ex. 7.6.

In Example 4.1 of Sect. 4.5, we considered a 1D random walk model of a polymer with n segments pointed in the positive x-direction and $N-m$ segments pointed in the negative x-direction. The total number of configurations was $\Omega(n) = N!/[n!(N-n)!]$, which contributed an entropic term $S(n) = k_B \ln \Omega(n)$ to the free energy. In this case the relevant quantity was the end-to-end distance $x = 2n - N$ and the conjugate entropic force was determined by TdS/dx. For the translocation problem shown in Fig. 7.20, there are two 3D chains, each of which is restricted to lie on one side of the membrane. Now we are interested in how the total number of configurations of each chain depends on m (the number of translocated segments). It can be shown that the total configurational entropy is [462, 615]

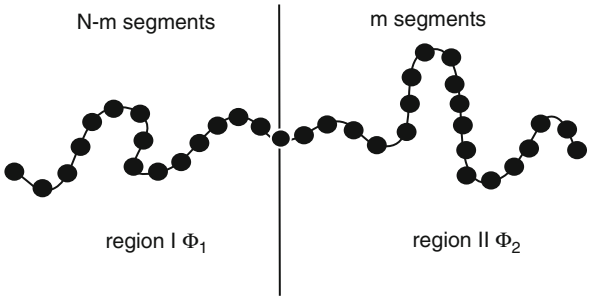


Fig. 7.20: Schematic illustration of a flexible biopolymer translating through a pore of a thin membrane. The electrical potential in each region is denoted by Φ_j, $j = 1, 2$

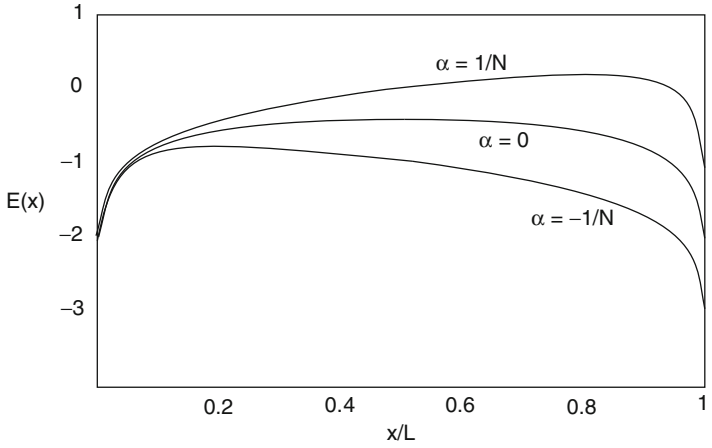


Fig. 7.21: Plot of free energy $E(x)$ (in units of k_BT) as a function of the length x of the right-hand segment with $\gamma = 0.69$ and $L = 1$. The three curves correspond to different values of the potential energy difference with $\alpha = \Delta\mu/k_BT$. (Redrawn from Muthukumar [462].)

$$S(m) = -k_B(1-\gamma)\left[\ln(m) + \ln(N-m)\right], \tag{7.3.29}$$

where $\gamma \approx 0.69$ is a constant that takes into account the fact that a 3D polymer cannot self-intersect (self-avoiding random walk); for a pure random walk $\gamma = 0.5$. A translocation step of the polymer leads to an increase in m, $m \to m+1$, which is opposed by an entropic force $f_{\text{ent}} = TdS/dm$ (assuming m is large so that it can be treated as a continuous variable). There is also an electrical field acting on the (charged) polymer due to the fact that there is a potential difference $V = \Phi_1 - \Phi_2$ across the membrane. If each segment has total charge $z_s e$, where e is the charge on an electron, then the change in potential energy when $m \to m+1$ is $\Delta\mu = ez_sV$. It follows that the free energy of the system for a given m is

$$E(m) = k_BT(1-\gamma)\left[\ln(m) + \ln(N-m)\right] + m\Delta\mu.$$

Setting $x = ma$ and $L = Na$, we can rewrite the free energy as

$$E(x) = k_BT(1-\gamma)\left[\ln(x/a) + \ln((L-x)/a)\right] + x\Delta\mu/a. \tag{7.3.30}$$

Plots of $E(x)$ as a function of x are shown in Fig. 7.21. It is now possible to reformulate translocation as Brownian motion through an energy barrier $E(x)$ with associated force $F(x) = -dE(x)/dx$ and diffusion coefficient D. One difference between various models is the assumed x-dependence of D [493]. Here we will follow [462] and treat D as a constant. The resulting FP equation is then

$$\frac{\partial p(x,t)}{\partial t} = D\frac{\partial}{\partial x}\left[\frac{p(x,t)}{k_BT}\frac{\partial E(x)}{\partial x} + \frac{\partial p(x,t)}{\partial x}\right]. \tag{7.3.31}$$

To complete the model, absorbing boundary conditions are introduced at $x = 0, L$,

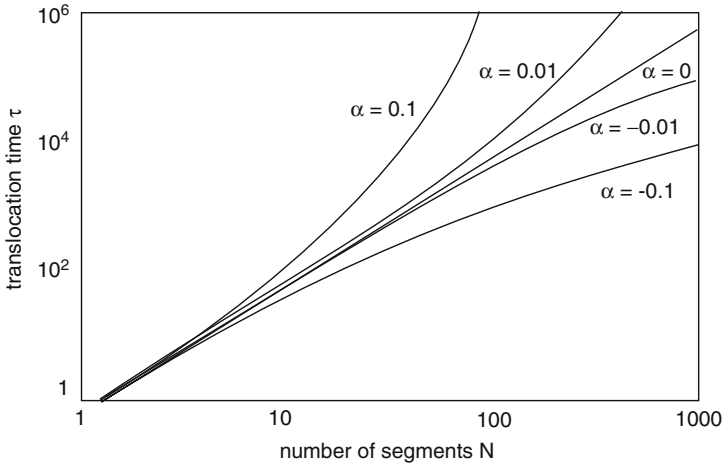


Fig. 7.22: Plot of MFPT τ (in units of a^2/D) for translocation of an N unit polymer and various values of $\alpha = \Delta\mu/k_BT$. (Redrawn from Muthukumar [462].)

$$p(0,t) = p(L,t) = 0, \tag{7.3.32}$$

under the assumption that if the polymer leaves the pore from either side, it never returns. It follows that the conditional MFPT τ for successful translocation, that is, the polymer crosses the boundary at $x = L$ rather than $x = 0$, can be calculated along the lines outlined in Sect. 2.3. One finds the following asymptotic results for large N [462] (see also Fig. 7.22):

$$\begin{aligned}
&\frac{D\tau}{a^2} \sim \frac{k_BT}{|\Delta\mu|}N, \quad \Delta\mu \ll 0,\\
&\frac{D\tau}{a^2} \sim \left(\frac{k_BT}{\Delta\mu}\right)^2 \exp\left(N\frac{\Delta\mu}{k_BT}\right), \quad \Delta\mu \gg 0,\\
&\frac{D\tau}{a^2} \sim N^2, \quad \Delta\mu = 0.
\end{aligned}$$

7.4 PDE Models of Active Transport

When considering the active transport of intracellular cargo over relatively long distances, it is often convenient to ignore the microscopic details of how a motor performs a single step (as described by the Brownian ratchet models of Sect. 4.3)

and to focus instead on the transitions between different types of motion (e.g., anterograde vs. retrograde active transport, diffusion vs. active transport). This has motivated a class of mesoscopic models that take the form of a system of PDEs [68, 197, 310, 367, 397, 473, 524, 601].

7.4.1 Active Transport on a 1D Track

For the sake of illustration, consider a simple three-state model of a particle moving on a 1D track of length L. Such a track could represent a single microtubular filament. Within the interior of the track, $0 < x < L$, the particle is taken to be in one of three states labeled by $n = 0, \pm$: unbound from the track and stationary (or slowly diffusing) ($n = 0$), bound to the track and moving to the right (anterograde) with speed v_+ ($n = +$), or bound to the track and moving to the left (retrograde) with speed $-v_-$ ($n = -$). For simplicity, take $v_\pm = v > 0$. Transitions between the three states are governed by a discrete Markov process. Let $Z(t)$ and $N(t)$ denote the random position and state of the particle at time t and define $\mathbb{P}(x,n,t \mid y,m,0)dx$ as the joint probability that $x \leq Z(t) < x + dx$ and $N(t) = n$ given that initially the particle was at position $Z(0) = y$ and was in state $N(0) = m$. Setting

$$p_n(x,t) \equiv \sum_m \mathbb{P}(x,t,n|0,0,m)\sigma_m \tag{7.4.1}$$

with initial condition $p_n(x,0) = \delta(x)\sigma_n$, $\sum_m \sigma_m = 1$, the evolution of the probability is described by the following system of PDEs for $t > 0$:

$$\frac{\partial p_+}{\partial t} = -v\partial_x p_+ - \beta_+ p_+ + \alpha p_0 \tag{7.4.2a}$$

$$\frac{\partial p_-}{\partial t} = v\partial_x p_- - \beta_- p_- + \alpha p_0 \tag{7.4.2b}$$

$$\frac{\partial p_0}{\partial t} = \beta_+ p_+ + \beta_- p_- - 2\alpha p_0. \tag{7.4.2c}$$

Here $\alpha, \beta_\pm$ are the transition rates between the stationary and mobile states. Equation (7.4.2) is supplemented by appropriate boundary condition at $x = 0, L$. For example, a reflecting boundary at $x = 0$ and an absorbing boundary at $x = L$ means that

$$p_-(0,t) = p_+(0,t), \quad p_-(L,t) = 0. \tag{7.4.3}$$

In the general case that the velocities $v_\pm$ in the two directions are different, the transport will be biased in the anterograde (retrograde) direction if $v_+/\beta_+ > v_-/\beta_-$ ($v_+/\beta_+ < v_-/\beta_-$). Note that the system of Eq. (7.4.2) is a three-state version of the Dogterom–Leibler model of microtubule catastrophe [146], Eq. (4.1.10), and the velocity-jump model of chemotaxis given by Eq. (5.3.14).

The three-state model is a special case of a general class of motor transport models, in which there are N distinct velocity states labeled by $n = 1,\ldots,N$ with corresponding velocities v_n. The probability density evolves according to the system of PDEs

$$\frac{\partial p}{\partial t} = -v_n \frac{\partial [p(x,n,t)]}{\partial x} + \sum_{n'=1}^{N} A(n,n';x)p(x,n',t), \tag{7.4.4}$$

where $A(n,n',x)$ specifies the transition rates between the N internal motor states and could, in principle, be space-dependent. We take $v_n > 0$ for $n = 1,\ldots,m$ and $v_n \leq 0$ for $n = m+1,\ldots,N$ with $m > 0$. For particular choices of A one recovers the three-state model and also the ToW model introduced in Sect. 4.4.2.

Example 7.1 (The three-state model). Let $n = 1,2,3$ denote the positive, negative, and stationary states, respectively. Then $A(n,m;x) = A_{nm}$ with the 3×3 matrix

$$\mathbf{A} = \begin{pmatrix} -\beta_+ & 0 & \alpha \\ 0 & -\beta_- & \alpha \\ \beta_+ & \beta_- & -2\alpha \end{pmatrix}.$$

Example 7.2 (ToW model). Consider a motor complex consisting of N_+ anterograde motors and N_- retrograde motors. The internal states of the complex are specified by the number of bound motors (n_+,n_-) and the velocity $v_c(n_+,n_-)$ of a given state is given by Eq. (4.4.13). Following [472, 473], we introduce the mapping $(n_+,n_-) \to n \equiv (N_+ + 1)n_- + (n_+ + 1)$ with $0 \leq n \leq N = (N_+ + 1)(N_- + 1)$. The corresponding probability density $p(n,x,t)$ satisfies Eq. (7.4.4) with $v_n = v_c(n_+,n_-)$. The components A_{nm}, $n,m = 1,\ldots,N$, of the state transition matrix A are given by the corresponding binding/unbinding rates of Eq. (4.4.9). That is, the nonzero off-diagonal terms are

$$\begin{aligned} A_{nm} &= \pi_+(n_+ - 1) \text{ for } m = n(n_+ - 1, n_-), \\ A_{nm} &= \pi_-(n_- - 1), \text{ for } m = n(n_+, n_- - 1), \\ A_{nm} &= \gamma_+(n_+ + 1), \text{ for } m = n(n_+ + 1, n_-), \\ A_{nm} &= \gamma_-(n_- + 1), \text{ for } m = n(n_+, n_- + 1). \end{aligned}$$

The diagonal terms are then given by $A_{nn} = -\sum_{m \neq n} A_{mn}$.

Note that Eq. (7.4.4) is another example of a differential Chapman–Kolmogrov (CK) equation for a stochastic hybrid system. That is, the dynamics combines a piecewise deterministic continuous process (in this case translation of the motor along the track) with a discrete stochastic process given by a continuous-time Markov process (in this case switching between different internal motors states). We have encountered other examples of stochastic hybrid systems in Chap. 3 (stochastic ion channels) and Chap. 6 (stochastic gene networks).

In many applications, one finds that the transition rates are fast compared to v/L where $v = \max_n |v_n|$. Performing the rescalings $x \to x/L$ and $t \to tv/L$ leads to a non-dimensionalized version of the CK equation

$$\frac{\partial p}{\partial t} = -v_n \frac{\partial [p(x,n,t)]}{\partial x} + \frac{1}{\varepsilon} \sum_{n'=1}^{N} A(n,n';x) p(x,n',t), \tag{7.4.5}$$

with $0 < \varepsilon \ll 1$.The transition matrix $A(n,m;x)$ is assumed to be irreducible for all x with a unique stationary density (right eigenvector) $\rho(x,n)$. In the limit $\varepsilon \to 0$, $p(x,n,t) \to \rho(x,n)$ and the motor moves deterministically according to the mean-field equation

$$\frac{dx}{dt} = V(x) \equiv \sum_{n=1}^{N} v_n \rho(x,n). \tag{7.4.6}$$

In the regime $0 < \varepsilon \ll 1$, there are typically a large number of transitions between different motor complex states n while the position x hardly changes at all. This suggests that the system rapidly converges to the (quasi) steady state $\rho(x,n)$, which will then be perturbed as x slowly evolves. The resulting perturbations can be analyzed using a QSS diffusion or adiabatic approximation, in which the CK equation (7.4.5) is approximated by a Fokker–Planck (FP) equation for the total density $C(x,t) = \sum_n p(x,n,t)$. The QSS approximation was first developed from a probabilistic perspective by Papanicolaou [494]; see also [204]. It has subsequently been applied to a wide range of problems in biology, including bacterial chemotaxis [140, 170, 262, 486, 487] (see also Sect. 5.3.3), wave-like behavior in models of slow axonal transport [196, 198, 524], and molecular motor-based models of random intermittent search [70, 472–474]. The QSS reduction proceeds as follows:

1. Decompose the probability density as

$$p(x,n,t) = C(x,t)\rho(x,n) + \varepsilon w(x,n,t), \tag{7.4.7}$$

where $\sum_n p(x,n,t) = C(x,t)$ and $\sum_n w(x,n,t) = 0$. Substituting into Eq. (7.4.5) yields

$$\frac{\partial C}{\partial t}\rho(x,n) + \varepsilon \frac{\partial w(x,n,t)}{\partial t} = -v_n \frac{\partial [C(x,t)\rho(x,n) + \varepsilon w(x,n,t)]}{\partial x} + \frac{1}{\varepsilon} \sum_{n'=1}^{N} A(n,n';x)[C(x,t)\rho(x,n') + \varepsilon w(x,n',t)].$$

Summing both sides with respect to n then gives

$$\frac{\partial C}{\partial t} = -\frac{\partial VC}{\partial x} - \varepsilon \sum_{n=1}^{N} v_n \frac{\partial w(x,n,t)}{\partial x}, \tag{7.4.8}$$

where $V(x) = \sum_m v_m \rho(x,m)$.

2. Using the equation for C and the fact that $A\rho = 0$, we have

$$\varepsilon \frac{\partial w}{\partial t} = \sum_{n'=1}^{N} A(n,n';x)w(x,n',t) - v_n \frac{\partial \rho(x,n)C}{\partial x} + \rho(x,n)\frac{\partial V(x)C}{\partial x} - \varepsilon \sum_{m=1}^{N} [v_n \delta_{m,n} - \rho(x,n)v_m] \frac{\partial w(x,m,t)}{\partial x}.$$

3. Introduce the asymptotic expansion

$$w \sim w_0 + \varepsilon w_1 + \varepsilon^2 w_2 + \ldots$$

and collect $O(1)$ terms:

$$\sum_{n'=1}^{N} A(n,n';x)w_0(n',x,t) = v_n \frac{\partial \rho(x,n)C(x,t)}{\partial x} - \rho(x,n)\frac{\partial V(x)C(x,t)}{\partial x}. \quad (7.4.9)$$

The Fredholm alternative theorem shows that this has a solution, which is unique on imposing the condition $\sum_n w_0(x,n,t) = 0$ (see Box 7B).

4. Combining Eqs. (7.4.8) and (7.4.9) shows that C evolves according to the FP equation

$$\frac{\partial C}{\partial t} = -\frac{\partial}{\partial x}(VC) + \varepsilon \frac{\partial}{\partial x}\left(D\frac{\partial C}{\partial x}\right) \quad (7.4.10)$$

with the drift V and diffusion coefficient D given by

$$V(x) = \sum_{m=1}^{N} v_m \rho(x,m), \quad D(x) = \sum_{n=1}^{N} Z(x,n)v_n, \quad (7.4.11)$$

where $Z(x,n)$ is the unique solution to

$$\sum_{m=1}^{N} A(n,m;x)Z(x,m) = [V(x) - v_n]\rho(x,n) \quad (7.4.12)$$

with $\sum_m Z(v,m) = 0$. We have dropped $O(\varepsilon)$ corrections to the drift term.

The FP equation (7.4.10) is often easier to analyze than the full CK equation, particularly when the number of internal states is large or the motor moves along a microtubular network rather than a single track. In the simple case of the three-state model given by Eq. (7.4.2), one can obtain explicit expressions for the drift and diffusion coefficient (see Ex. 7.7):

$$V = \frac{v}{\gamma}\left(\frac{1}{\beta_+} - \frac{1}{\beta_-}\right), \quad D = \varepsilon\left(\frac{(1-V)^2}{\gamma\beta_+^2} + \frac{(1+V)^2}{\gamma\beta_-^2}\right), \quad \gamma = \frac{1}{\beta_+} + \frac{1}{\beta_-} + \frac{1}{\alpha}.$$

The QSS approximation will be used to study the efficacy of motor intracellular cargo transport in Sect. 7.6.

Box 7B. Fredholm alternative theorem (matrices).

Consider an M-dimensional linear inhomogeneous equation $\mathbf{Az} = \mathbf{b}$ with $\mathbf{z}, \mathbf{b} \in \mathbb{R}^M$. Suppose that the $M \times M$ matrix $\mathbf{A}$ has a nontrivial null-space and let $\mathbf{u}$ be a null vector of the adjoint matrix $\mathbf{A}^\dagger$, that is, $\mathbf{A}^\dagger \mathbf{u} = 0$. The Fredholm alternative theorem for finite-dimensional vector spaces states that the inhomogeneous equation has a (nonunique) solution for $\mathbf{z}$ if and only if $\mathbf{u} \cdot \mathbf{b} = 0$ for all null vectors $\mathbf{u}$.

Let us apply this theorem to Eq. (7.4.9) for fixed x, t with $A_{nm} = A(n,m;x)$ and $A^\dagger_{nm} = A_{mn}$. The one-dimensional null-space is spanned by the vector with components $u_n = 1$, since $\sum_n u_n A_{nm} = \sum_n A^\dagger_{mn} u_n = 0$. Hence Eq. (7.4.9) has a solution provided that

$$0 = \sum_n \left[v_n \frac{\partial \rho(x,n) C(x,t)}{\partial x} - \rho(x,n) \frac{\partial V(x) C(x,t)}{\partial x} \right].$$

This immediately follows since $\sum_n \rho(x,n) = 1$ and $\sum_n \rho(x,n) v_n = V(x)$ for all x.

7.4.2 Active Transport on Microtubular Networks

In the case of axonal or dendritic transport in neurons, the microtubles tend to be aligned in parallel so that one can treat the transport process as effectively 1D. On the other hand, intracellular transport within the soma of neurons and most nonpolarized animal cells occurs along a microtubular network that projects radially from organizing centers (centrosome) with outward polarity [3] (see Sect. 7.3). This allows the delivery of cargo to and from the nucleus. Moreover, various animal viruses including HIV take advantage of microtubule-based transport in order to reach the nucleus from the cell surface and release their genome through nuclear pores [369]. In contrast, the delivery of cargo from the cell membrane or nucleus to other localized cellular compartments requires a nonradial path involving several tracks. It has also been found that microtubules bend due to large internal stresses, resulting in a locally disordered network. This suggests that in vivo transport on relatively short length scales may be similar to transport observed in vitro, where microtubular networks are not grown from centrosomes, and thus exhibit orientational and polarity disorder [313, 553]. Another example where a disordered microtubular network exists is within the *Drosophila* oocyte [26]. Kinesin and dynein motor-driven transport along this network is thought to be one of the mechanisms for establishing the asymmetric localization of four maternal mRNAs—*gurken, oskar, bicoid*, and *nanos*—which are essential for the development of the embryonic body axes.

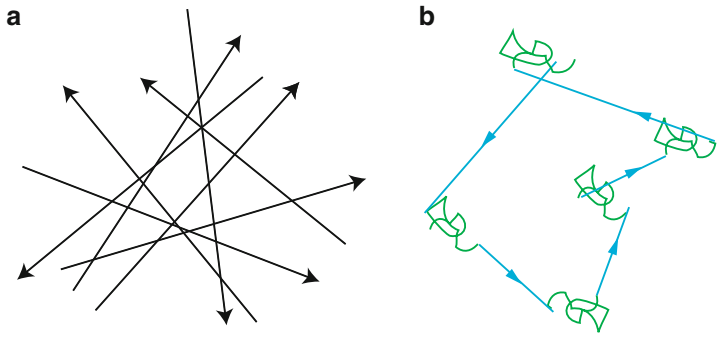


Fig. 7.23: Active transport on a disordered microtubular network. (**a**) Random orientational arrangement of microtubles. (**b**) Effective 2D random intermittent search in which a particle switches between diffusion and ballistic motion in a random direction

A detailed microscopic model of intracellular transport within the cell would need to specify the spatial distribution of microtubular orientations and polarity, in order to determine which velocity states are available to a motor–cargo complex at a particular spatial location. However, a simplified model can be obtained under the "homogenization" assumption that the network is sufficiently dense so that the set of velocity states (and associated state transitions) available to a motor complex is independent of position. In that case, one can effectively represent the active transport and delivery of cargo to an unknown target within the cell in terms of a two- or three-dimensional model of active transport [31, 32, 397].

For simplicity, consider a disordered 2D microtubular network as illustrated in Fig. 7.23. (The extension to 3D networks is relatively straightforward.) Suppose that after homogenization, a molecular motor at any point $\mathbf{r} = (x,y)$ in the plane can bind to a microtubule with any orientation θ, resulting in ballistic motion with velocity $\mathbf{v}(\theta) = v(\cos\theta, \sin\theta)$ and $\theta \in [0, 2\pi)$. If the motor is unbound, then it acts as a Brownian particle with diffusion coefficient D_0. Transitions between the diffusing state and a ballistic state are governed by a discrete Markov process. The transition rate β from a ballistic state with velocity $\mathbf{v}(\theta)$ to the diffusive state is taken to be independent of θ, whereas the reverse transition rate is taken to be of the form $\alpha q(\theta)$ with $\int_0^{2\pi} q(\theta)d\theta = 1$. Suppose that at time t the motor is undergoing ballistic motion. Let $(X(t),Y(t))$ be the current position of the motor particle and let $\Theta(t)$ denote the corresponding velocity direction. Introduce the conditional probability density $p(x,y,\theta,t)$ such that $p(x,y,\theta,t)dxdyd\theta$ is the joint probability that $(x,y,\theta) < (X(t),Y(t),\Theta(t)) < (x+dx, y+dy, \theta+d\theta)$ given that the particle is in the ballistic phase. Similarly, take $p_0(x,y,t)$ to be the corresponding conditional probability density if the particle is in the diffusive phase. (For the moment the initial conditions are left unspecified.) The evolution of the probability densities for $t > 0$ can then be described in terms of the following 2D system of PDEs [70]:

$$\frac{\partial p}{\partial t} = -\nabla \cdot (\mathbf{v}(\theta)p) - \frac{\beta}{\varepsilon} p(\mathbf{r},\theta,t) + \frac{\alpha q(\theta)}{\varepsilon} p_0(\mathbf{r},t) \tag{7.4.13a}$$

$$\frac{\partial p_0}{\partial t} = \varepsilon D_0 \nabla^2 p_0 + \frac{\beta}{\varepsilon} \int_0^{2\pi} p(\mathbf{r},\theta',t) d\theta' - \frac{\alpha}{\varepsilon} p_0(\mathbf{r},t). \tag{7.4.13b}$$

In the case of a uniform density, $q(\theta) = 1/(2\pi)$, Eqs. (7.4.13a) and (7.4.13b) reduce to a 2D model of active transport considered by Benichou et al. [31, 32, 397]. The units of space and time have been fixed according to $l = 1$ and $l/v = 1$, where l is a typical run length. Furthermore, for the given choice of units, it has been assumed that there exists a small parameter $\varepsilon \ll 1$ such that all transition rates are $O(\varepsilon^{-1})$, the diffusivity is $O(\varepsilon)$, and all velocities are $O(1)$. In the limit $\varepsilon \to 0$, the system rapidly converges to the space-clamped (i.e., $\nabla p = \nabla p_0 = 0$) steady-state distributions $(p^{\rm ss}(\theta), p_0^{\rm ss})$ where

$$p_0^{\rm ss} = \frac{\beta}{\alpha+\beta} \equiv b, \quad p^{\rm ss}(\theta) = \frac{\alpha q(\theta)}{\alpha+\beta} \equiv a q(\theta). \tag{7.4.14}$$

In the regime $0 < \varepsilon \ll 1$, it is possible to extend the QSS approximation to the 2D system to obtain a multivariate FP equation.

As in the 1D case, the QSS approximation is based on the assumption that for $0 < \varepsilon \ll 1$, solutions remain close to the steady-state solution. Hence,

$$p(\mathbf{r},\theta,t) = u(\mathbf{r},t) p^{\rm ss}(\theta) + \varepsilon w(\mathbf{r},\theta,t) \tag{7.4.15a}$$

$$p_0(\mathbf{r},t) = u(\mathbf{r},t) p_0^{\rm ss} + \varepsilon w_0(\mathbf{r},t), \tag{7.4.15b}$$

where

$$u(\mathbf{r},t) \equiv \int_0^{2\pi} p(\mathbf{r},\theta,t) d\theta + p_0(\mathbf{r},t)$$

and

$$\int_0^{2\pi} w(\mathbf{r},\theta,t) d\theta + w_0(\mathbf{r},t) = 0. \tag{7.4.16}$$

Furthermore, the initial conditions are taken to be

$$u(\mathbf{r},0) = \delta(\mathbf{r}-\mathbf{X}), \quad w(\mathbf{r},0) = w_0(\mathbf{r},0) = 0,$$

which are equivalent to the following initial conditions for the full probability densities:

$$p(\mathbf{r},\theta,0) = \delta(\mathbf{r}-\mathbf{X}) p^{\rm ss}(\theta),\ p_0(\mathbf{r},0) = \delta(\mathbf{r}-\mathbf{X}) p_0^{\rm ss}.$$

Thus, the initial internal state of the motor (diffusive or ballistic with velocity $\mathbf{v}(\theta)$) is generated according to the steady-state distributions $p^{ss}(\theta)$ and p_0^{ss}. In other words, the motor starts on the slow manifold of the underlying dynamics. If this were not the case, then one would need to carry out a multi-scale analysis in order to take into account the initial transient dynamics transverse to the slow manifold [204]. Perturbation and projection methods can now be used to derive a closed

equation for the scalar component $u(\mathbf{r},t)$ [70]. First, integrating Eq. (7.4.13a) with respect to θ and adding Eq. (7.4.13b) yields

$$\begin{aligned}\frac{\partial u}{\partial t} &= \varepsilon D_0 \nabla^2 p_0 - \langle\langle \mathbf{v}\cdot\nabla p\rangle\rangle \\ &= \varepsilon b D_0 \nabla^2 u - a\langle \mathbf{v}\rangle\cdot\nabla u - \varepsilon\langle\langle \mathbf{v}\cdot\nabla w\rangle\rangle + O(\varepsilon^2),\end{aligned} \tag{7.4.17}$$

where $\langle f\rangle = \int_0^{2\pi} q(\theta) f(\theta) d\theta$ and $\langle\langle f\rangle\rangle = \int_0^{2\pi} f(\theta) d\theta$ for any function or vector component $f(\theta)$. Next, substituting Eqs. (7.4.15a) and (7.4.15b) into Eqs. (7.4.13a) and (7.4.13b) yields

$$\begin{aligned}aq(\theta)\frac{\partial u}{\partial t} + \varepsilon\frac{\partial w}{\partial t} &= -\mathbf{v}(\theta)\cdot\nabla\left[aq(\theta)u + \varepsilon w\right] - \beta w \\ &\quad + \alpha q(\theta) w_0.\end{aligned} \tag{7.4.18}$$

and

$$\begin{aligned}b\frac{\partial u}{\partial t} + \varepsilon\frac{\partial w_0}{\partial t} &= \varepsilon D_0\nabla^2\left(bu + \varepsilon w_0\right) \\ &\quad + \beta\langle w\rangle - \alpha w_0.\end{aligned} \tag{7.4.19}$$

Now substitute Eq. (7.4.17) into Eqs. (7.4.18) and (7.4.19). Collecting terms to leading order in ε and using Eq. (7.4.16) then gives

$$w_0(\mathbf{r},t) \sim \frac{ab}{\alpha+\beta}\left[\langle\mathbf{v}\rangle\cdot\nabla u\right], \tag{7.4.20}$$

and

$$w(\mathbf{r},\theta,t) \sim \frac{q(\theta)}{\beta}\left(a^2(1+b)\langle\mathbf{v}\rangle - a\mathbf{v}(\theta)\right)\cdot\nabla u. \tag{7.4.21}$$

Finally, substituting Eqs. (7.4.21) and (7.4.20) into Eq. (7.4.17) yields to $O(\varepsilon)$ the FP equation

$$\frac{\partial u}{\partial t} = -\nabla\cdot(\mathbf{V}u) + \varepsilon b D_0\nabla^2 u + \varepsilon\nabla\cdot(\mathbf{D}\nabla u). \tag{7.4.22}$$

The diffusion tensor $\mathbf{D}$ has components $D_{kl}, k = x,y, l = x,y$

$$D_{kl} \sim \frac{a}{\beta}\left(\langle v_k v_l\rangle - \langle v_k\rangle\langle v_l\rangle + b^2\langle v_k\rangle\langle v_l\rangle\right), \tag{7.4.23}$$

to lowest order in ε, while the effective drift velocity is given by $\mathbf{V} \sim a\langle\mathbf{v}\rangle$. Here $\langle v_n\rangle = \int_0^{2\pi} v_n(\theta) q(\theta) d\theta$.

In the case of a uniform direction distribution $q(\theta) = 1/(2\pi)$, the diffusion tensor reduces to a scalar. This follows from the fact that $v_x = v\cos\theta, v_y = v\sin\theta$ so $\langle v_x\rangle = \langle v_y\rangle = \langle v_x v_y\rangle = 0$ and to leading order

$$D_{xx} = \frac{av^2}{2\beta} = D_{yy}, \quad D_{xy} = 0. \tag{7.4.24}$$

More generally, assuming that $\mathscr{Q}(\theta)$ is sufficiently smooth, we can expand it as a Fourier series,

$$\mathscr{Q}(\theta) = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{n=1}^{\infty}(\omega_n \cos(n\theta) + \hat{\omega}_n \sin(n\theta)). \tag{7.4.25}$$

Assume further that $\omega_1 = \hat{\omega}_1 = 0$ so there is no velocity bias, i.e., $\langle v_x \rangle = \langle v_y \rangle = 0$. Then

$$\begin{aligned} D_{xx} &= \frac{av^2}{\beta}\int_0^{2\pi} \cos^2(\theta)\mathscr{Q}(\theta)d\theta = \frac{av^2}{2\beta}(1+\omega_2) \\ D_{yy} &= \frac{av^2}{\beta}\int_0^{2\pi} \sin^2(\theta)\mathscr{Q}(\theta)d\theta = \frac{av^2}{2\beta}(1-\omega_2), \\ D_{xy} &= \frac{av^2}{\beta}\int_0^{2\pi} \sin(\theta)\cos(\theta)\mathscr{Q}(\theta)d\theta = \frac{av^2}{2\beta}\hat{\omega}_2. \end{aligned} \tag{7.4.26}$$

It follows that only the second terms in the Fourier series expansion contribute to the diffusion tensor.

An alternative formulation of transport on disordered microtubular networks has been developed by Kahana et al. [313] in terms of random velocity fields [522, 706]. In order to describe the basic idea, consider the simplified model analyzed by [706]. The latter model consists of a set of equally spaced parallel tracks along the x-axis, for example (see Fig. 7.24). The tracks are assigned random polarities ± 1 with equal probabilities corresponding to quenched polarity disorder. A particle undergoes a random walk in the y-direction, whereas when a particle attaches to a certain track it moves ballistically with velocity ± 1 according to the track's polarity. It is assumed that when a particle hops to a neighboring track it binds immediately. Let $X(t)$ denote the displacement of a random walker in the longitudinal direction at time t:

$$X(t) = \int_0^t v[y(t')]dt'. \tag{7.4.27}$$

Taking the continuum limit in the y-direction means that

$$p(y,t) = \frac{1}{\sqrt{4\pi Dt}} \mathrm{e}^{-y^2/4Dt},$$

where D is the diffusion coefficient, and the velocity field is δ-correlated $\langle v(y)v(y')\rangle_c = v^2\xi\,\delta(y-y')$. Here averaging is taken with respect to the quenched polarity disorder and ξ is the infinitesimal spacing between tracks. Now consider the second moment $\langle\langle X^2(t)\rangle\rangle$ of the stochastic process averaged with respect to the quenched disorder and realizations of the random walk:

$$\langle\langle X^2(t)\rangle\rangle = 2\int_0^t dt_1 \int_0^{t_1} dt_2 \langle\langle v[y(t_1)]v[y(t_2)]\rangle\rangle, \tag{7.4.28}$$

where

$$\langle\langle v[y(t_1)]v[y(t_2)]\rangle\rangle = \int_{-\infty}^{\infty} dy_1 \int_{-\infty}^{\infty} dy_2 \langle v(y_1)v(y_2)\rangle_c \times p(y_2,t_2)p(y_1-y_2,t_1-t_2). \tag{7.4.29}$$

Using Laplace transforms and the velocity correlation function,

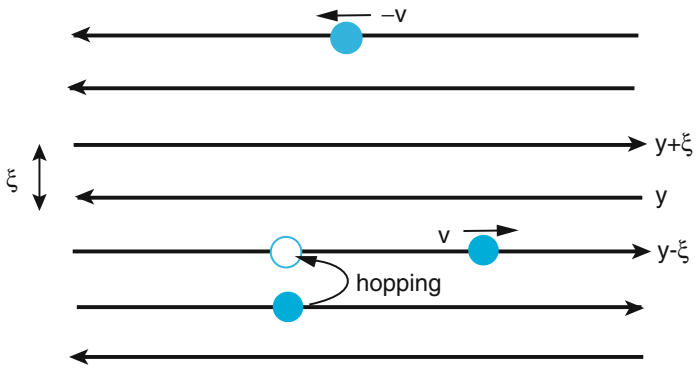


Fig. 7.24: Random velocity model of a microtubular network with quenched polarity disorder. Particles move ballistically along parallel tracks in a direction determined by the polarity of the given track. They also hop between tracks according to an unbiased random walk

$$\langle\langle \tilde{X}^2(s)\rangle\rangle = \frac{2v^2\xi}{s}\tilde{p}(0,s)\int_{-\infty}^{\infty}\tilde{p}(y,s)dy, \tag{7.4.30}$$

with

$$\tilde{p}(y,s) = \frac{1}{\sqrt{4Ds}}\mathrm{e}^{-|y|\sqrt{s/D}}.$$

Performing the integration with respect to y thus shows that $\langle\langle \widetilde{X^2}(s)\rangle\rangle = v^2\xi D^{-1/2}s^{-5/2}$, which on inverting the Laplace transform gives

$$\langle\langle X^2(t)\rangle\rangle = \frac{4v^2\xi}{3\sqrt{\pi D}}t^{3/2}. \tag{7.4.31}$$

An equivalent formulation of the problem is to treat $\langle\langle X^2(t)\rangle\rangle$ as the solution to the differential equation [313]

$$\frac{d^2}{dt^2}\langle\langle X^2(t)\rangle\rangle = 2v^2\xi y p(0,t), \tag{7.4.32}$$

where $\xi p(0,t)$ is the probability of turn to the origin at time t within a single lattice spacing ξ and $p(0,t) = 1/\sqrt{4\pi Dt}$. In conclusion, the random velocity model supports anomalous superdiffusion in the x-direction.

Kahana et al. [313] extended the above construction to 2D (and 3D) disordered networks where there are parallel tracks in the x- and y-directions. The distribution of polarities are unbiased in both directions. A self-consistent description of the dynamics is obtained by taking

$$\frac{d^2}{dt^2}\langle\langle X^2(t)\rangle\rangle = 2v^2\xi p_y(0,t),\ \frac{d^2}{dt^2}\langle\langle Y^2(t)\rangle\rangle = 2v^2\xi p_x(0,t), \tag{7.4.33}$$

where p_x and p_y are the probability densities of the x- and y-coordinates. From the symmetry of the network, $p_x(0,t) = p_y(0,t)$. Hence, assuming that $p_x(0,t) = C\langle\langle X^2(t)\rangle\rangle^{-1/2}$ for some constant C and setting $\phi(t) = \langle\langle X^2(t)\rangle\rangle$ gives

$$\phi^{1/2}\frac{d^2}{dt^2}\phi = 2Cv^2\xi. \tag{7.4.34}$$

It follows that $\phi(t) \sim t^{4/3}$ so that the diffusion is less enhanced than in the case of parallel tracks in one direction. Finally, note that active transport on the randomly oriented network of Fig. 7.23 exhibits normal rather than anomalous diffusion. A major difference from the random velocity model is that the latter has quenched polarity disorder, whereas the former has dynamical polarity disorder.

7.5 Exclusion Processes

So far we have considered a single molecular motor or motor/cargo complex moving along a filament track. However, in practice there could be many active particles moving along the same track, which could interact with each other and exhibit some form of collective behavior. This has motivated a number of studies that model the movement of multiple motor particles as an ASEP [2, 176, 339, 344, 481, 496, 497, 513, 515]. In the simplest version of such models, each particle hops unidirectionally at a uniform rate along a 1D lattice; the only interaction between particles is a hard-core repulsion that prevents more than one particle occupying the same lattice site at the same time. This so-called totally asymmetric exclusion process (TASEP) is combined with absorption/desorption kinetics, in which individual particles can bind to or unbind from the track (see Fig. 7.25). The TASEP has become the paradigmatic model of nonequilibrium stochastic processes, and a variety of analytical methods have been developed to generate exact solutions for the stationary state (see [52, 115, 567] and references therein). However, when chemical kinetic or other biologically motivated extensions of TASEP are included, it is usually no longer possible to obtain exact solutions so that some form of mean-field approximation is required.

7.5.1 Asymmetric Exclusion Process and the Hydrodynamic Limit

Let us consider in more detail the system shown in Fig. 7.25, which consists of a finite 1D lattice of N sites labeled $i = 1,\ldots,N$. The microscopic state of the system is given by the configuration $\mathscr{C}$ that specifies the distribution of identical particles on the lattice. That is, $\mathscr{C} = \{n_1,\ldots,n_N\}$ where each occupation number $n_i = 1$ if the ith site is occupied by a single particle and $n_i = 0$ if the site is vacant. Exclusion effects preclude more than one particle at any site. Thus, the state space consists of 2^N configurations. Let $\mathscr{P}(\mathscr{C},t)$ demote the probability of finding a particular configuration $\mathscr{C}$ at time t. The evolution of this probability distribution is described by a master equation:

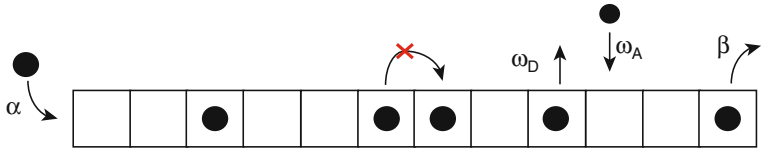


Fig. 7.25: Schematic diagram of TASEP with absorption/desorption kinetics, in which particles can spontaneously detach and attach at rates ω_D and ω_A, respectively

$$\frac{d\mathscr{P}(\mathscr{C},t)}{dt} = \sum_{\mathscr{C}'\neq\mathscr{C}} \left[\mathscr{W}_{\mathscr{C}'\to\mathscr{C}}\mathscr{P}(\mathscr{C}',t) - \mathscr{W}_{\mathscr{C}\to\mathscr{C}'}\mathscr{P}(\mathscr{C},t)\right]. \tag{7.5.1}$$

The transition rate $\mathscr{W}_{\mathscr{C}\to\mathscr{C}'}$ from configuration $\mathscr{C}$ to $\mathscr{C}'$ is determined from the following set of rules [496]:

(a) at sites $i = 1,\ldots,N-1$, a particle can jump to site $i+1$ at a unit rate if the latter is unoccupied;
(b) at site $i = 1$ ($i = N$) a particle can enter (exit) the lattice at a rate α (β) provided that the site is unoccupied (occupied);
(c) in the bulk of the lattice, a particle can detach from a site at a rate ω_D and attach to an unoccupied site at a rate ω_A.

Rules (a) and (b) constitute a TASEP with open boundary conditions, whereas rule (c) describes absorption/desorption kinetics. It follows that the evolution of the particle densities $\langle n_i\rangle$ away from the boundaries is given by the exact equation

$$\frac{d\langle n_i\rangle}{dt} = \langle n_{i-1}(1-n_i)\rangle - \langle n_i(1-n_{i+1})\rangle + \omega_A\langle 1-n_i\rangle - \omega_D\langle n_i\rangle. \tag{7.5.2}$$

Here $\langle n_i(t)\rangle = \sum_{\mathscr{C}} n_i \mathscr{P}(\mathscr{C},t)$. Similarly, at the boundaries

$$\frac{d\langle n_1\rangle}{dt} = -\langle n_1(1-n_2)\rangle + \alpha\langle 1-n_1\rangle - \omega_D\langle n_1\rangle, \tag{7.5.3a}$$

$$\frac{d\langle n_N\rangle}{dt} = \langle n_{N-1}(1-n_N)\rangle + \omega_A\langle 1-n_N\rangle - \beta\langle n_N\rangle. \tag{7.5.3b}$$

Note that in the absence of any exclusion constraints, Eq. (7.5.2) reduces to a spatially discrete version of the unidirectional PDE (7.4.2a), with $p_+(n_i\Delta x,t) = \langle n_i\rangle$, $\beta_+ = \omega_D$, and $p_0\alpha = \omega_A$ and $v_+/\Delta x = 1$. The goal is to find a nonequilibrium stationary state for which the current flux along the lattice is a constant J_0. It then follows that J_0 has the exact form

$$J_0 = \alpha\langle 1-n_1\rangle = \langle n_i(1-n_{i+1})\rangle = \beta\langle n_N\rangle, \quad i = 1, N-1.$$

Equations (7.5.2) and (7.5.3) constitute a nontrivial many-body problem, since in order to calculate the time evolution of $\langle n_i\rangle$ it is necessary to know the two-point correlations $\langle n_{i-1}(1-n_i)\rangle$. The latter obey dynamical equations involving three-point and four-point correlations. Thus, there is an infinite hierarchy of equations of motion. However, progress can be made by using a mean-field approximation and a continuum limit in order to derive a PDE for the density of particles [176, 497]. The mean-field approximation consists of replacing two-point correlations by products of single-site averages:

$$\langle n_i n_j\rangle = \langle n_i\rangle\langle n_j\rangle.$$

(It turns out that for pure TASEP (no binding or unbinding) this yields an accurate phase diagram; see below.) Next introduce the infinitesimal lattice spacing ε and set $x = k\varepsilon$, $\rho(x,t) = \rho_k(t) \equiv \langle n_k(t)\rangle$. The continuum limit is then defined according to $N \to \infty$ and $\varepsilon \to 0$ such that the length of the track $L = N\varepsilon$ is fixed. (Fix length scales by setting $L = 1$.) Expanding $\rho_{k\pm1}(t) = \rho(x \pm \varepsilon, t)$ in powers of ε gives

$$\rho(x\pm\varepsilon,t) = \rho(x) \pm \varepsilon\partial_x\rho(x,t) + \frac{1}{2}\varepsilon^2\partial_{xx}\rho(x,t) + O(\varepsilon^3).$$

Finally, rescaling the absorption/desorption rates according to $\omega_A = \Omega_A\varepsilon, \omega_D = \Omega_D\varepsilon$, and rescaling time $\tau = \varepsilon t$, Eq. (7.5.2) becomes to $O(\varepsilon)$

$$\frac{\partial\rho}{\partial\tau} = \frac{\varepsilon}{2}\frac{\partial^2\rho}{\partial x^2} - (1-2\rho)\frac{\partial\rho}{\partial x} + \Omega_A(1-\rho) - \Omega_D\rho. \tag{7.5.4}$$

Similarly, Eq. (7.5.3) reduces to the boundary conditions

$$J(0,t) = \alpha(1-\rho(0,t)), \quad J(L,t) = \beta\rho(L,t)$$

where the continuum flux is

$$J(x,t) = -\frac{\varepsilon}{2}\frac{\partial\rho}{\partial x} + \rho(1-\rho). \tag{7.5.5}$$

In the following, we describe methods for analyzing the mean-field model for a pure TASEP by setting $\Omega_A = \Omega_D = 0$. For extensions to the full model, we refer the reader to [176, 496, 497].

7.5.2 Steady-State Analysis

In order to develop the basic theory, we will focus on pure TASEP by setting $\Omega_A = \Omega_D = 0$ in Eq. (7.5.4). We proceed by finding a stationary nonequilibrium state for which the current $J(x,t) = J_0$ is constant and determining the corresponding stationary density profile. This then generates a phase diagram with respect to the parameters α, β, which can be calculated explicitly [52, 355]. The steady-state current equation takes the form

$$\varepsilon \frac{d\rho}{dx} = \rho(1-\rho) - J_0.$$

Setting $q = \rho - 1/2$, this becomes

$$\varepsilon \frac{dq}{dx} = v^2 - q^2, \quad v^2 = \frac{1}{4} - J_0.$$

It follows that for $v^2 > 0$

$$\varepsilon \int \frac{dq}{(v-q)(v+q)} = x - x_0,$$

where x_0 is an integration constant. Using partial fractions, we find that

$$\frac{v+q}{v-q} = \mathrm{e}^{2v(x-x_0)/\varepsilon},$$

which on rearranging yields the density profile

$$\rho(x) = \frac{1}{2} + v \tanh(v(x-x_0)/\varepsilon), \tag{7.5.6}$$

with $v \geq 0$. On the other hand, if $v^2 < 0$, then we have

$$\varepsilon \int \frac{dq}{|v^2| + q^2} = x - x_0.$$

Under the change of variables $q = \cot \mathrm{an}(u)$, we can evaluate the integral and find that

$$\rho(x) = 0.5 + |v| \mathrm{cotan}(|v|(x-x_0)/\varepsilon). \tag{7.5.7}$$

The two unknown parameters J_0, x_0 can be determined in terms of α, β by imposing the boundary conditions at $x = 0, L$. The resulting phase diagram in the limit of large L and small ε is shown in Fig. 7.26. Three distinct phases can be identified:

1. A *low-density* phase in which the bulk density is smaller than $1/2$, $x_0 = O(L)$ and $v^2 > 0$. Since $\varepsilon \ll 1$, we see from Eq. (7.5.6) that $\rho(x) \approx 0.5 - v$ for all $x < x_0$. In particular, at the left-hand boundary $\alpha(0.5 + v) = J_0$, which can be rewritten as $v = J_0/\alpha - 0.5$. Squaring both sides and using the definition of v gives, to lowest order in ε,
$$\rho(0) = \alpha, \quad J_0 = \alpha(1-\alpha), \quad \alpha < 1/2.$$
The other boundary condition becomes
$$\beta = \frac{J_0}{0.5 + v\tanh(v(L-x_0)/\varepsilon)} > \frac{J_0}{0.5+v} = \alpha.$$

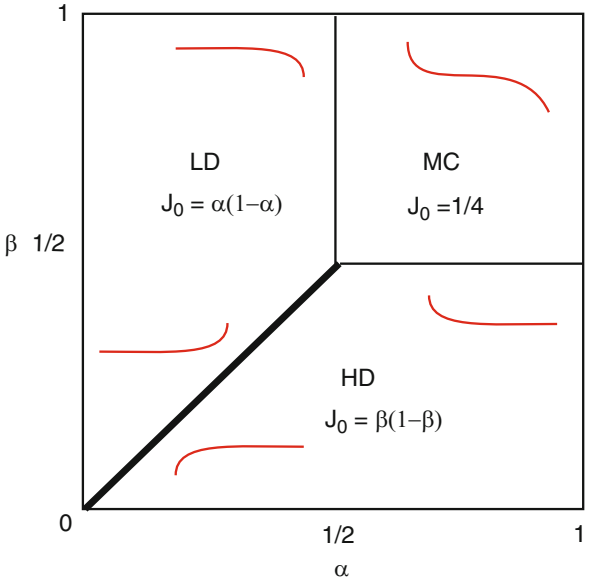


Fig. 7.26: Mean-field phase diagram for the TASEP showing the regions of α, β parameter space where the low-density (*LD*), high-density (*HD*), and maximal-current (*MC*) phases exist. Schematic illustrations of the density profiles in the various regions are shown in *red*

In order to satisfy this boundary condition, there is an ε-wide boundary layer at $x = L$ with $L - x_0 = O(\varepsilon)$.

2. A *high-density* phase in which the bulk density is larger than $1/2$ and $x_0 \approx 0$. Hence, $\rho(x) \approx 0.5 + v$ in the bulk of the domain and at the right-hand boundary

we have $\beta(0.5+v) = J_0$. Following along similar lines to the low-density case, we deduce that

$$\rho(L) = 1-\beta, \quad J_0 = \beta(1-\beta), \quad \beta < 1/2,$$

and $\beta < \alpha$. There is now a boundary layer around $x = 0$ in order to match the rate α. The two phases coexist along the line $\alpha = \beta < 1/2$.

3. A *maximal-current phase*. In the region $\alpha > 1/2, \beta > 1/2$, we require $J_0 > 1/4$ so that $v^2 < 0$. It turns out that the current takes the form $J_0 = 0.5 + O(\varepsilon^2/L^2)$, that is, it is very close to the maximal value of function $\rho(1-\rho)$. This follows from the observation that the solution (7.5.7) will blow up unless $0 < |v|(x-x_0)/\varepsilon < \pi$ for all $x \in [0,L]$. This implies that $x_0 = -O(\varepsilon)$ and $|v| < \pi\varepsilon/L$. Under these conditions, Eq. (7.5.7) ensures that $\rho(x) \approx 0.5$ in the bulk of the domain. The precise values of v and x_0 are then adjusted so that the boundary conditions at $x = 0, L$ are satisfied: $\rho(0) = 1 - 1/(4\alpha) > 0.5$ and $\rho(L) = 1/(4\beta) < 0.5$. Also note away from the left-hand boundary, we have $\text{cotan}(|v|(x-x_0)/\varepsilon) \approx \varepsilon/(|v|x)$ so that

$$\rho(x) \sim 0.5 + \varepsilon/x.$$

7.5.3 Method of Characteristics and Shocks

Equation (7.5.4) is mathematically similar in form to the viscous Burger's equation with additional source terms [554]. Thus, one expects singularities such as shocks in the density ρ to develop in the inviscid or non-dissipative limit $\varepsilon \to 0^+$. One can view the formation and propagation of shocks as a way of understanding how the system evolves to the final steady-state solution [176, 349]. Again, we will illustrate this by considering a pure TASEP. Setting $\Omega_A = \Omega_D = 0$ and $\varepsilon = 0$ in Eq. (7.5.4), yields a kinematic wave equation of the quasilinear form

$$\frac{\partial \rho}{\partial \tau} + \frac{\partial J(\rho)}{\partial x} = 0, \quad J(\rho) = \rho(1-\rho). \tag{7.5.8}$$

Equation (7.5.8) is a particular example of a quasilinear PDE [see Eq. (10.1.6)], and can be analyzed using the method of characteristics introduced in Sect. 3.6. Thus one looks for characteristic curves $x = x(\tau)$ along which $\rho(\tau) \equiv \rho(x(\tau), \tau)$ satisfies

$$\frac{d\rho}{d\tau} = \frac{\partial \rho}{\partial \tau} + \frac{dx}{d\tau}\frac{\partial \rho}{\partial x}.$$

Comparison with Eq. (7.5.8) leads to the characteristic equations

$$\frac{dx}{d\tau} = J'(\rho) = 1 - 2\rho, \quad \frac{d\rho}{d\tau} = 0. \tag{7.5.9}$$

It can be seen that the characteristics are straight lines along which ρ is constant. Suppose that $x(0) = x_0$ and the corresponding initial density is $\rho(x_0, 0) = \rho_0(x_0)$. For simplicity, we ignore the boundary conditions by taking $x \in \mathbb{R}$. The corresponding characteristic solutions (parameterized by x_0) are then

$$x(\tau) = [1 - 2\rho_0(x_0)]t + x_0, \quad \rho(\tau) = \rho(x(\tau), \tau) = \rho_0(x_0).$$

In other words, the density profile at time t is determined by the propagation of the initial density $\rho(x_0, 0)$ along the straight line characteristics.

For the given kinetic wave equation, one finds that an initial density profile can sharpen up to form a discontinuity, which then propagates as a shock wave. This is illustrated in Fig. 7.27 for an initial density given by the piecewise linear function $\rho_0(x) = 0$ for $x < 0$, $\rho_0(x) = x$ for $0 \le x \le 1$, and $\rho_0(x) = 1$ for $x > 1$. Since higher densities propagate more slowly than lower densities, an initial linear density profile steepens until a shock is formed at the points of intersection where pairs of characteristics meet. In general, a shock propagates with a speed v_S determined by the so-called Rankine–Hugonoit condition [554]:

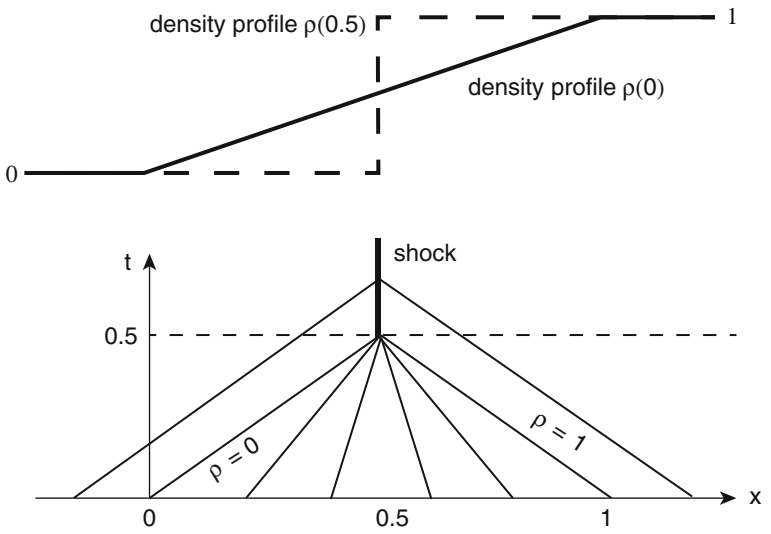


Fig. 7.27: Formation of a shock for Eq. (7.5.8). The characteristics are straight lines of speed $1 - 2\rho$ with ρ constant along a characteristic. The initial density profile evolves into a stationary shock solution

$$v_S = \frac{J(\rho_2) - J(\rho_1)}{\rho_2 - \rho_1} = 1 - \rho_1 - \rho_2 \tag{7.5.10}$$

where ρ_1, ρ_2 are the densities on either side of the shock. For the particular initial density profile shown in Fig. 7.27, $\rho_1 = 0$ and $\rho_2 = 1$ so that the shock is stationary ($v_S = 0$). The possibility of stationary shocks reflects the fact that the current

$J(\rho) = \rho(1-\rho)$ has a maximum, which means that two different densities can have the same current on either side of the shock. The Rankine–Hugonoit condition is usually derived by considering weak solutions of the kinematic equation (see Box 7C). However, it can also be understood in terms of a traveling wave solution of the corresponding PDE with weak diffusion (see Ex. 7.8). A further example illustrating the method of characteristics and shocks is given in Ex. 7.9.

Box 7C. Weak formulation of shocks and the Rankine–Hugonoit condition.

In order to deal with discontinuities and shocks, it is necessary to introduce a more flexible notion of a solution to the kinematic wave equation, in which derivatives of the solution are not directly involved [554]. Let $\phi(x,t)$ be a smooth function in $\mathbb{R} \times [0,\infty)$ with compact support, that is, it vanishes outside a bounded domain. If ρ is a smooth solution of the kinematic wave equation $\partial_t \rho + \partial_x J(\rho) = 0$ with initial condition $\rho(x,0) = p(x)$, then

$$\int_0^\infty \int_{\mathbb{R}} [\partial_t \rho + \partial_x J(\rho)] \phi \, dx \, dt = 0.$$

We now carry out an integration by parts of the first term with respect to t and the second term with respect to x:

$$\int_0^\infty \partial_t \rho \, \phi \, dx dt = -\int_0^\infty \rho \, \partial_t \phi \, dx dt - \int_{\mathbb{R}} p(x) \phi(x,0) dx$$

and, since $\phi(\pm\infty) = 0$,

$$\int_0^\infty \partial_x J(\rho) \phi \, dx dt = -\int_0^\infty J(\rho) \partial_x \phi \, dx dt.$$

We thus obtain the integral equation

$$\int_0^\infty \int_{\mathbb{R}} [\rho \, \partial_t \phi + J(\rho) \partial_x \phi] dx dt + \int_{\mathbb{R}} p(x) \phi(x,0) dx = 0. \qquad (7.5.11)$$

It can be seen that no derivative of ρ appears. We define a *weak solution* of the kinematic wave equation to be one that satisfies the integral equation (7.5.11) for every test function ϕ in $\mathbb{R} \times [0,\infty)$ with compact support. Note that if ρ is also smooth, then we can reverse the integration by parts to recover the PDE.

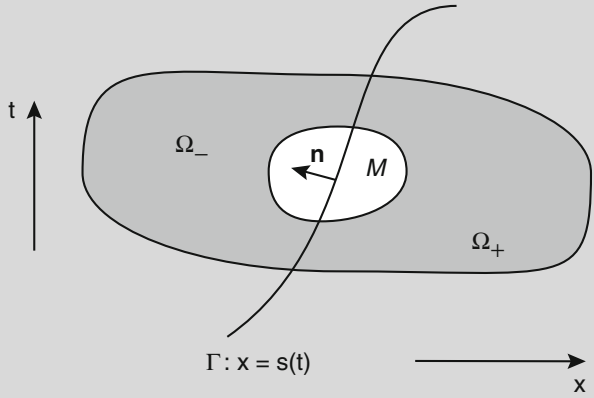


Fig. 7.28: Construction of a weak solution in a domain partitioned by a shock Γ

The weak formulation can now be used to derive the Rankine–Hugonoit condition for the speed of a shock. Suppose that an open set $\Omega \in \mathbb{R} \times [0,\infty)$is partitioned into two disjoint domains $\Omega_{\pm}$ by a smooth shock curve Γ satisfying $x = s(t)$. Suppose that ρ is a weak solution in Ω that is a continuously differentiable function ρ in the closed sets $\overline{\Omega_+}$ and $\overline{\Omega_-}$. That is, ρ is a smooth solution of $\partial_t \rho + \partial_x J(\rho) = 0$ in Ω_+ and Ω_- such that ρ and its first derivatives extend continuously up to Γ from either side (see Fig. 7.28). Choose a test function ϕ with support in a compact set $M \subset \Omega$ such that $M \cap \Omega$ is not empty, and take $\phi(x,0) = 0$. The integral equation for the weak solution takes the form

$$\begin{aligned} 0 &= \int_0^\infty \int_{\mathbb{R}} [\rho \partial_t \phi + J(\rho)\partial_x \phi] dx dt \\ &= \int_{\Omega_+} [\rho \partial_t \phi + J(\rho)\partial_x \phi] dx dt + \int_{\Omega_-} [\rho \partial_t \phi + J(\rho)\partial_x \phi] dx dt. \end{aligned}$$

Integrating by parts the integral over Γ_+ using Stoke's Theorem, noting that $\phi = 0$ on $\partial\Omega_+/\Gamma$ (the boundary of Ω_+ excluding the curve Γ), we have

$$\begin{aligned} \int_{\Omega_+} [\rho \partial_t \phi + J(\rho)\partial_x \phi] dx dt &= -\int_{\Omega_+} [\partial_t \rho + \partial_x J(\rho)] \phi dx dt \\ &\quad + \int_\Gamma [\rho_+ n_2 + J(\rho_+) n_1] \phi dl \\ &= \int_\Gamma [\rho_+ n_2 + J(\rho_+) n_1] \phi dl, \end{aligned}$$

where ρ_+ denotes the value of ρ on Γ from the Ω_+ side, $\mathbf{n} = (n_1, n_2)$ is the outward normal vector on the boundary $\partial\Omega_+$, and dl denotes arc length along Γ. Similarly,

$$\int_{\Omega_-} [\rho \partial_t \phi + J(\rho)\partial_x \phi] dx\, dt = -\int_\Gamma [\rho_- n_2 + J(\rho_-) n_1] \phi\, dl,$$

where ρ_- denotes the value of ρ on Γ from the Ω_- side.

The above analysis shows that

$$\int_\Gamma [(J(\rho_+) - J(\rho_-)) n_1 + (\rho_+ - \rho_-) n_2]\, \phi\, dl = 0.$$

The arbitrariness of ϕ means that

$$(J(\rho_+) - J(\rho_-)) n_1 + (\rho_+ - \rho_-) n_2 = 0. \tag{7.5.12}$$

If ρ were continuous across Γ, then this equation would be an identity. Therefore, suppose that $\rho_+ \neq \rho_-$. The shock curve is given by $x = s(t)$, which implies that

$$\mathbf{n} = (n_1, n_2) = \frac{1}{\sqrt{1 + \dot{s}(t)^2}} (-1, \dot{s}(t)).$$

Substituting for $\mathbf{n}$ in Eq. (7.5.12) thus gives the Rankine–Hugonoit condition for the speed $\dot{s}(t)$ of the shock:

$$\dot{s} = \frac{J(\rho_+(s,t)) - J(\rho_-(s,t))}{\rho_+(s,t) - \rho_-(s,t)}. \tag{7.5.13}$$

The method of characteristics and kinematic wave theory yields insights into the dynamics underlying the formation of the various stationary phases shown in Fig. 7.26 [52, 349, 355]. The basic idea is to consider kinematic waves propagating from the left-hand and right-hand boundaries, respectively, by considering an initial density profile such that $\rho(0,0) = \alpha$ and $\rho(L,0) = 1 - \beta$ with L large.

1. If $\alpha, \beta < 1/2$, then a kinematic wave propagates from the left-hand and right-hand boundaries with speeds $1 - 2\alpha > 0$ and $(2\beta - 1) < 0$, respectively. These waves thus propagate into the interior of the domain and meet somewhere in the middle to form a shock that propagates with speed $v_S = \beta - \alpha$. If $\beta > \alpha$, then the shock moves to the right-hand boundary and the bulk of the domain is in a low-density (LD) state with $\rho \approx \alpha < 1/2$. On the other hand, if $\beta < \alpha$, then the shock moves to the left-hand boundary and the bulk of the domain is in a high-density (HD) state with $\rho \approx 1 - \beta > 1/2$. For weak dissipation the sharp drop in the density at one end is smoothed to form a boundary layer.

2. In the special case $\alpha = \beta < 1/2$ the LD and HD phases coexist. The solution consists of a low-density region separated from a high-density region by a shock. Once higher-order dissipative effects are included, this shock diffuses freely between the ends of the domain, so that the average density profile is linear.
3. If both $\alpha > 1/2$ and $\beta > 1/2$, then the steady-state bulk solution has the maximal-current density $J = 1/4$. In order to show this, and to determine how bulk solutions match the boundary conditions, it is necessary to include dissipation effects as in the previous section.

The above analysis based on the theory of shocks can be extended to the full molecular motor model that combines TASEP with binding/unbinding kinetics [176, 496, 497]. When $\Omega_A, \Omega_D \neq 0$ the characteristic Eq. (7.5.9) become

$$\frac{dx}{d\tau} = 1 - 2\rho, \quad \frac{d\rho}{d\tau} = \Omega_A(1-\rho) - \Omega_D\rho. \tag{7.5.14}$$

It follows that the characteristics are now curves in the $x-t$ plane. For example, consider the propagation of density fluctuations along a characteristic starting at the left boundary with $\rho = \alpha < 1/2$ and $\alpha < K/(K+1)$, where $K = \Omega_A/\Omega_D$. It follows from Eq. (7.5.14) that initially the fluctuation propagates along the characteristic with decreasing speed and increasing density. If $K/(1+K) < 1/2$, then ρ will approach the constant value $\rho = K/(K+1)$ and the speed approaches a constant value. However, if $K/(1+K) > 1/2$, then after a finite time the density reaches $\rho = 1/2$ and propagation ceases. A similar analysis holds for characteristics propagating from the right boundary. Furthermore, characteristics propagating from opposite boundaries can again intersect, implying multivalued densities and the formation of shocks. The resulting shock has the same wave speed as pure TASEP. Of particular interest are

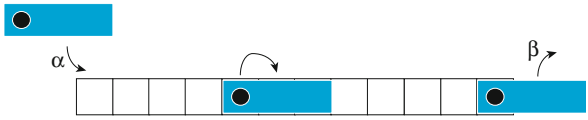


Fig. 7.29: A TASEP with extended particles of size $l = 3$

stationary solutions for which the current $J = \rho(1-\rho)$ is constant so that any shock solution is stationary ($v_S = 0$). To a first approximation, these can be obtained by finding steady-state solutions of the mean-field equation

$$(1-2\rho)\frac{\partial\rho}{\partial x} - \Omega_D[K - (1+K)\rho] = 0. \tag{7.5.15}$$

The occurrence of stationary shocks is consistent with the observation that this is a first-order ordinary differential equation (ODE), but there are two boundary conditions. One thus proceeds by integrating from the left boundary where $\rho(0) = \alpha$ to obtain a density profile $\rho_L(x)$ and then integrating from the right

boundary where $\rho(L) = 1 - \beta$ to obtain a second density profile $\rho_R(x)$. The full solution is constructed by matching the two profiles at a shock whose position also has to be determined. If the shock lies outside the interval $[0,L]$, then it is necessary to include at least one boundary layer. A detailed analysis of the steady-state solutions with coexisting low- and high-density phases and the corresponding phase diagram with respect to the parameters $(\alpha,\beta,\Omega_D,\Omega_A)$ can be found in [176, 497]. If the effects of dissipation are also taken into account, then the sharp interfaces and boundary layers become smooth fronts of size $O(1/\varepsilon)$.

One of the first examples of a TASEP model in biology was proposed by Gibbs and collaborators in their study of the translation of mRNA by ribosomes during protein synthesis [404, 405] (see also Sect. 6.7.1). However, it is necessary to modify pure TASEP to include multisite particles, since ribosomes are large molecules which extend over several codons or lattice sites (around $l = 12$). In the case of multisite particles, one has to specify the rules for entry and exit of a ribosome [115, 703]. One possibility is "complete entry, incremental exit," which assumes that a ribosome enters completely provided the first l lattice sites are vacant, whereas it exits one step at a time [113] (see Fig. 7.29). Inclusion of extended objects considerably complicates the analysis even though the basic structure of the phase diagram is preserved [113, 584]. In contrast to pure TASEP, there does not currently exist an exact solution, although mean-field approximations do provide useful insights. A second biologically motivated modification of TASEP is to include site-dependent hopping rates [114, 149, 190, 344]. This is motivated by the fact that the local hopping rate depends on the relative abundance of specific amino-acid carrying tRNA. Using a combination of Monte Carlo simulations and mean-field theory it can be shown, for example, that two defects (regions of slow hopping rates) decrease the steady-state current more when they are close to each other. Finally, note that more complex models take into account intermediate steps in the translocation of a ribosome along the mRNA, including the binding of tRNA to the ribosome and hydrolysis [25, 118, 202, 528].

7.6 Random Intermittent Search Processes

Random search strategies are used throughout nature as a means of efficiently searching for one or more targets of unknown location. Examples include animals foraging for food or shelter [28, 659, 660], proteins searching for particular sites on DNA [42, 250], and biochemical reaction kinetics [397]. Recently, there has been a great deal of interest in a particular class of search strategy known as random intermittent search, in which a particle randomly switches between a slow search phase and a faster non-search phase [30–32, 485]. At the macroscopic scale, intermittent motion has been observed in a variety of animal species during exploratory behavior. One striking example is given by the nematode *C. elegans*, which alternates between a fast displacement along a straight trajectory (roaming) and a much slower displacement along a more sinuous trajectory (dwelling) [510]. During the slow phase,

the worm's head, bearing most of its sensory organs, moves and touches the surface nearby, suggesting that this is a search phase. Intermittent motion also occurs at the microscopic level of reaction kinetics within biological cells (see Fig. 7.30). For example, diffusing molecules within the cellular environment intermittently bind and unbind to molecular motors that execute ballistic transport along cytoskeletal filaments powered by ATP hydrolysis; reaction with a target molecule can only occur during the free diffusive phase [288, 397]. The effective reaction rate can be determined by solving a FPT problem for random intermittent search in 3D or 2D, depending on whether or not the search domain is restricted to the plasma membrane. An effective 1D search process can occur in tubular structures such as the axons and dendrites of neurons, where microtubular filaments tend to align in parallel. Experimental observations of active transport along axons and dendrite reveal intermittent behavior with constant velocity movement in both directions along a microtubule, interrupted by brief pauses or fast oscillatory movements that may correspond to localization at specific targets such as synapses [153, 342, 544].

A variety of stochastic models of random intermittent search processes have been developed (as reviewed in [32, 397]). In these studies it is typically assumed that (i) a particle (e.g., a motor–cargo complex) is searching for some hidden target (e.g., a subcellular compartment) within a bounded physical domain (e.g., the plasma membrane or cytoplasm), (ii) the motion of the particle is unbiased, (iii) the particle initiates its search at some random location within the domain, and (iv) the probability of eventually finding the target is equal to unity. Under these conditions, it can be shown that there exists an optimal search strategy, in the sense that there exist values for the durations of each phase that minimize the mean search time to find a single hidden target [29–31]. An analogous result holds for protein–DNA interactions. However, for some cellular processes, such as the directed transport of newly synthesized products from the nucleus to targets in the plasma membrane, assumptions (ii)–(iv) no longer hold, since the motion is biased in the anterograde direction and the initial location is always at the nucleus. Moreover, there is now a nonzero probability that the particle does not reach the target due to degradation or absorption by another target [69, 471]. Under these circumstances, an optional search strategy no longer exists. On the other hand, the failure to find a subcellular target may be mitigated if the target is only partially hidden, in the sense that it emits a local chemical signal that increases the probability of a motor particle stopping in a neighborhood of that target [473].

7.6.1 Diffusive Search for a Protein-DNA Binding Site

A wide range of cellular processes are initiated by a protein transcription factor (see Sect. 6.1) binding a specific target sequence of base pairs (target site) on a long DNA molecule. The precise mechanism whereby a protein finds its DNA binding site remains unclear. However, it has been observed experimentally that reactions occur at very high rates, of around $k = 10^{10}\,\mathrm{M}^{-1}\,\mathrm{s}^{-1}$ [537, 540]. This is around 100 times

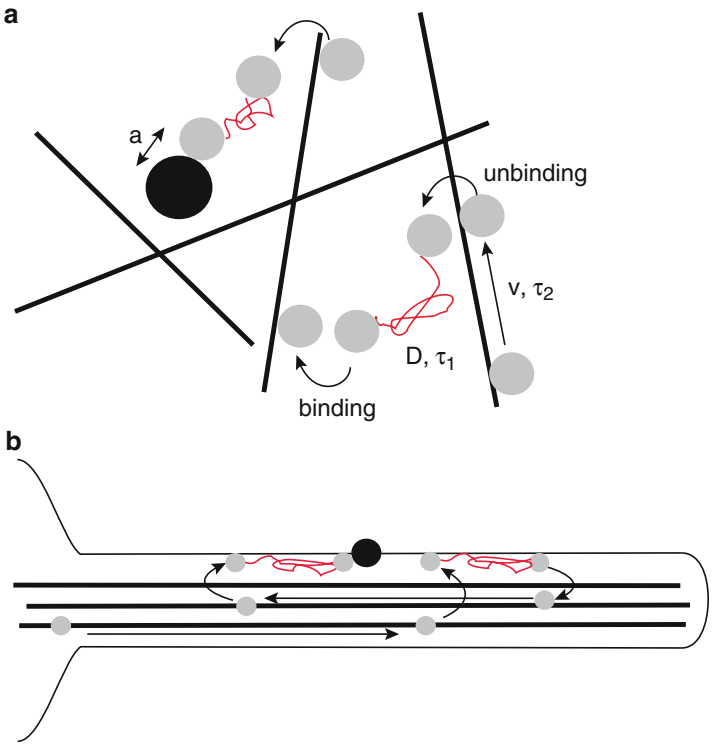


Fig. 7.30: Models of intracellular intermittent active transport in 3D and 1D. Redrawn from [397]. (**a**) A molecule alternates between a diffusive phase of mean duration τ_1 and diffusion coefficient D, and ballistic phases of velocity v and mean duration τ_2 that are powered by molecular motor transport along cytoskeletal filaments. The cytoskeletal filaments are in a disordered state (see also Fig. 7.23). If the molecule diffuses within a neighborhood of a target substrate, it can undergo a chemical reaction. (**b**) Active transport of intracellular cargo along a tubular structure such as an axon or dendrite of a neuron

faster than the rate based on the Smoluchowski theory of diffusion-limited reaction rates (Sect. 2.4) and 1,000 times higher than most known protein–protein association rates. This apparent discrepancy in reaction rates suggests that some form of facilitated diffusion occurs. The best known theoretical model of facilitated diffusion for proteins searching for DNA targets was originally developed by Berg, Winter, and von Hippel (BHW) [41, 42, 684] and subsequently extended by a number of groups [127, 250, 281, 440]. The basic idea of the BHW model is to assume that the protein randomly switches between two distinct phases of motion, 3D diffusion in solution and 1D diffusion along DNA (sliding) (see Fig. 7.31a). Such a mechanism is one example of a random intermittent search process. The BHW model assumes that there are no correlations between the two transport phases, so that the main factor in speeding up the search is an effective reduction in the dimensionality of the protein motion. However, as recently reviewed in [346], there are a number of discrepancies

between the BHW model and experimental data, which has led to several alternative theoretical approaches to facilitated diffusion. We first review the BHW model and then briefly discuss these alternative models.

A simple method for estimating the effective reaction rate of facilitated diffusion in the BHW model is as follows [440]. Consider a single protein searching for a single binding site on a long DNA strand of N base pairs, each of which has length b. Suppose that on a given search, there are R rounds labeled $i = 1, \ldots, R$. In the ith round the protein spends a time $T_{3,i}$ diffusing in the cytosol followed by a period $T_{1,i}$ sliding along the DNA. The total search time is thus $T = \sum_{i=1}^{R}(T_{3,i} + T_{1,i})$, and the mean search time is $\tau = r(\tau_3 + \tau_1)$. Here r is the mean number of rounds and τ_3, τ_1 are the mean durations of each phase of 3D and 1D diffusion. Let n denote the mean number of sites scanned during each sliding phase with $n \ll N$. If the binding site of DNA following a 3D diffusion phase is distributed uniformly along the DNA, then the probability of finding the specific promoter site is $p = n/N$. It follows that the probability of finding the site after R rounds is $(1-p)^{R-1}p$. Hence, the mean number of rounds is $r = 1/p = N/n$. Assuming that 1D sliding occurs via normal diffusion, then $nb = 2\sqrt{D_1\tau_1}$ where D_1 is the $1D$ diffusion coefficient, and we have [440]

$$\tau = \frac{N}{n}(\tau_1 + \tau_3). \tag{7.6.1}$$

Since τ_3 depends primarily on the cellular environment and is thus unlikely to vary significantly between proteins, it is erasable to minimize the mean search time with respect to τ_1 while τ_3 is kept fixed. Setting $d\tau/d\tau_1 = 0$ implies that the optimal search time occurs when $\tau_1 = \tau_3$ with $\tau_{\text{opt}} = 2N\tau_3/n$. On the other hand, the expected search time for pure 3D diffusion gives $\tau_{3D} = N\tau_3$, which is the approximate time to find one out of N sites by randomly binding to a single site of DNA every τ_3 seconds and no sliding ($\tau_1 = 0$). Thus facilitated diffusion is faster by a factor $n/2$.

Further insights to facilitated diffusion may be obtained by using the Smoluchowski formula for the rate at which a diffusing protein can find any one of N binding sites of size b, namely, $\tau_3^{-1} = 4\pi D_3 Nb[DNA]$, where $[DNA]$ is the concentration of DNA. (We are simplifying the problem by not worrying about the 3D geometry of DNA.) Using this to eliminate N shows that the effective reaction rate of facilitated diffusion is [440]

$$k \equiv \frac{1}{\tau[DNA]} = 4\pi D_3 \left(\frac{\tau_3}{\tau_1 + \tau_3}\right) nb$$

This equation identifies two competing mechanisms in facilitated diffusion. First, sliding diffusion effectively increases the reaction cross section from 1 to n base pairs, thus accelerating the search process compared to standard Smoluchowski theory. This is also known as the *antenna effect* [281]. However, the search is also slowed down by a factor $\tau_3/(\tau_1 + \tau_3)$, which is the fraction of the time the protein spends in solution. That is, a certain amount of time is lost by binding to nonspecific sites that are far from the target. Note that typical experimental values are

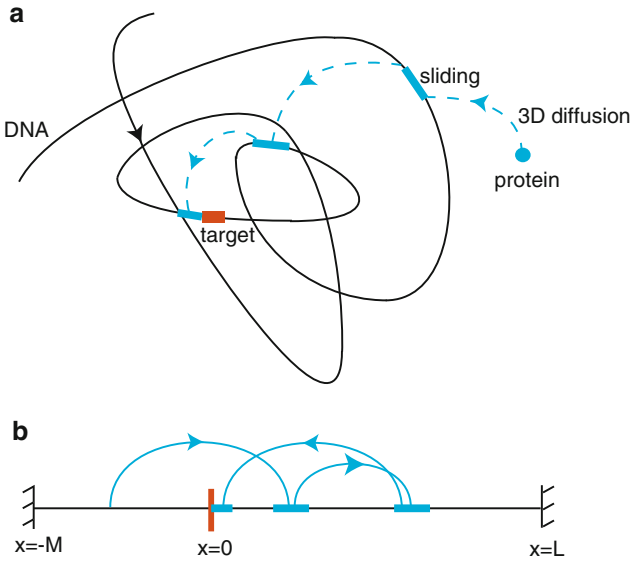


Fig. 7.31: (**a**) Mechanism of facilitated diffusion involving alternating phases of 3D diffusion and 1D diffusion (sliding along the DNA). (**b**) 1D representation of facilitated diffusion

$D_3 = 10\,\mu\text{m}^2\,\text{s}^{-1}$, $b = 0.34\,\text{nm}$, and $n = 200$, and one has to convert k into units of inverse molar per second.

A more complicated analysis is needed in order to take into account the effects of boundaries, for example. Here we review the particular formulation of Coppey et al. [127], which generalizes the original analysis of Berg [42]. Suppose that DNA is treated as a finite track of length $l = L + M$ with reflecting boundaries at $x = -M$ and $x = +L$ and a point-like target at $x = 0$ (see Fig. 7.31). Rather than modeling 3D diffusion explicitly, each time the protein dissociates from DNA it simply rebinds at a random site at a time t later that is generated from an exponential waiting time density. This is based on the assumption that 3D excursions are uncorrelated in space. It might be expected that excursions would be correlated due to the geometric configuration of DNA. However, in solution DNA is a random coil so that even short 3D trips can generate long displacements relative to the linear position of the protein along the DNA strand, resulting in decorrelation of excursions. If $P_3(t)$ denotes the probability density that the protein in solution at time $t = 0$ binds to the DNA at time t at a random position, then

$$P_3(t) = \lambda_3 \mathrm{e}^{-\lambda_3 t}, \tag{7.6.2}$$

where $\tau_3 = 1/\lambda_3$ is again the mean time spent in solution. Next, let $P_1(x,t)$ be the conditional probability density that the protein dissociates from the DNA at time t

without finding the target, given that it is at linear position x along the DNA at time $t = 0$:

$$P_1(x,t) = \lambda_1 \mathrm{e}^{-\lambda_1 t}\mathbb{P}(x,t) \tag{7.6.3}$$

where $\tau_1 = 1/\lambda_1$ is the mean time of each sliding phase and $\mathbb{P}(x,t)$ is the conditional probability density that the protein starting at x has not met the target at time t. Finally, let $Q_1(x,t)$ be the conditional probability density that the protein starting at x finds the target at time t:

$$Q_1(x,t) = \mathrm{e}^{-\lambda_1 t} f(x,t), \tag{7.6.4}$$

where $f(x,t) = -d\mathbb{P}(x,t)/dt$ is the FPT density associated with diffusion along the DNA strand. That is, $f(x,t)dt$ is the probability that starting at x at $t = 0$, the protein finds the target during a single phase of sliding diffusion in the time interval $[t, t+dt]$. (Protein–DNA binding is assumed to be diffusion-limited so that as soon as the protein reaches the target site it reacts.)

Suppose that in a given trial, a protein starting at x at time $t = 0$ executes $n-1$ excursions before finding the target with $t_1, \ldots, t_n$ the residence times on DNA and $\tau_1, \ldots, \tau_{n-1}$ the excursion times. The probability density for such a sequence of events with $t = \sum_{i=1}^{n} t_i + \sum_{i=1}^{n-1} \tau_n$ is

$$P_n(x, \{t_i, \tau_i\}) = Q_1(t_n)P_3(\tau_{n-1})P_1(t_{n-1}) \ldots P_1(t_2)P_3(\tau_1)P_1(x,t_1), \tag{7.6.5}$$

where $P_1(t) = \langle P_1(x,t)\rangle, Q_1(t) = \langle Q_1(x,t)\rangle$ and $\langle g(x,t)\rangle \equiv (L+M)^{-1}\int_{-M}^{L} g(x,t)dx$ for an arbitrary function g. In order to determine the FPT density $F(x,t)$ for finding the target, it is necessary to sum over all possible numbers of excursions and intervals of time, given the constraint $t = \sum_{i=1}^{n} t_i + \sum_{i=1}^{n-1} \tau_n$. Thus, setting $F(t) = \langle F(x,t)\rangle$, one finds that

$$F(t) = \sum_{n=1}^{\infty} \int_0^{\infty} dt_1 \ldots dt_n d\tau_1 \ldots d\tau_{n-1} \tag{7.6.6}$$
$$\delta\left(\sum_{i=1}^{n} t_i + \sum_{i=1}^{n-1} \tau_i - t\right) Q_1(t_n) \prod_{i=1}^{n-1} P_3(\tau_i) \prod_{i=1}^{n-1} P_1(t_i).$$

Finally, Laplace transforming this equation and using the convolution Theorem 2.1 gives [127]

$$\tilde{F}(s) = \tilde{f}(\lambda_1 + s)\left[1 - \frac{1-\tilde{f}(\lambda_1+s)}{(1+s/\lambda_1)(1+s/\lambda_3)}\right]^{-1}, \tag{7.6.7}$$

with $\tilde{f}(s) = \int_0^{\infty} \mathrm{e}^{-st}\langle f(x,t)\rangle dt$. Given $\tilde{F}(s)$, the MFPT to find the target (averaged over the starting position x) is then

$$\tau = \int_0^{\infty} tF(t)dt = -\left.\frac{d\tilde{F}(s)}{ds}\right|_{s=0}, \tag{7.6.8}$$

which can be evaluated to give

$$\tau = \frac{1-\tilde{f}(\lambda_1)}{\tilde{f}(\lambda_1)}(\lambda_1^{-1}+\lambda_3^{-1}). \tag{7.6.9}$$

All that remains is to determine $\tilde{f}(x,s)$ averaged with respect to x. If $x<0$ ($x>0$), then one simply needs to determine the FPT density for a 1D Brownian particle on the interval $[-M,0]$ ($[0,L]$) with a reflecting boundary at $x=-M$ ($x=L$) and an absorbing boundary at $x=0$. Recall from Sect. 2.3 that $f(x,t)$ satisfies the backward FP equation

$$\frac{\partial f(x,t)}{\partial t} = D_1 \frac{\partial^2 f(x,t)}{\partial x^2}, \tag{7.6.10}$$

with $f(x,0)=0$, $f(0,t)=\delta(t)$, and $\partial_x f(L,t)=0$ or $\partial_x f(-M,t)=0$. Taking Laplace transforms,

$$s\tilde{f}(x,s) = D_1 \frac{\partial^2 \tilde{f}(x,s)}{\partial x^2}, \tag{7.6.11}$$

with $\tilde{f}(0,s)=1, \partial_x \tilde{f}(L,s)=0$ or $\partial_x \tilde{f}(-M,s)=0$. The general solution is $\tilde{f}(x,s) = Ae^{-\sqrt{s/D_1}x} + Be^{-\sqrt{s/D_1}x}$ with the coefficients A,B determined by the boundary conditions. Solving for A,B separately when $x<0$ and $x>0$ and averaging with respect to x finally gives

$$\tilde{f}(s) = \frac{1}{L+M}\sqrt{\frac{D_1}{s}}\left[\tanh(L\sqrt{s/D_1}) + \tanh(M\sqrt{s/D_1})\right]. \tag{7.6.12}$$

Thus, setting $\tau_i = 1/\lambda_i$, $i=1,3$,

$$\tau = \left[\frac{(L+M)/\sqrt{\tau_1 D_1}}{\tanh(L/\sqrt{\tau_1 D_1}) + \tanh(M/\sqrt{\tau_1 D_1})} - 1\right](\tau_1+\tau_3), \tag{7.6.13}$$

which recovers the original result of Berg [42]. It also recovers Eq. (7.6.1) when $L/\sqrt{\tau_1 D_1}, M/\sqrt{\tau_1 D_1} \gg 1$. (Ex. 7.10 involves carrying out the missing steps in the above derivations.) A sample plot of the mean search time is shown in Fig. 7.32.

There have been a number of extensions of the BHW model that incorporate various biophysical effects. For example, sequence-dependent protein–DNA interactions generate a rugged energy landscape during sliding motion of the protein [440]. This observation then leads to an interesting speed–stability paradox [440, 585]. On the one hand, fast 1D search requires that the variance σ^2 of the protein–DNA binding energy be sufficiently small, that is, $\sigma \sim k_B T$, whereas stability of the protein at the DNA target site requires $\sigma \sim 5k_B T$. One suggested resolution of this paradox is to assume that a protein–DNA complex has two conformational states: a recognition state with large σ and a search state with small σ [440]. If the transitions between the states are sufficiently fast, then target stability and fast search can be reconciled. (For a recent review of the speed–stability paradox and its implications for search mechanisms see [585].) Other effects include changes in the conformational state of

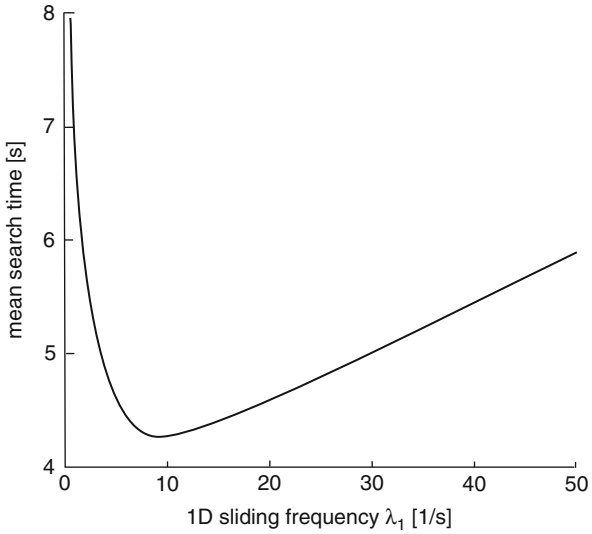


Fig. 7.32: The mean search time plotted against the 1D sliding frequency λ_1. The length of DNA is taken to be 5,000 bp, the 3D frequency $\lambda_3 = 10\,\mathrm{s}^{-1}$, and the 1D diffusion coefficient $D_1 = 5 \times 10^3\ \mathrm{bp}^2/\mathrm{s}$. (Redrawn from Coppey et al. [127].)

DNA and the possibility of correlated association/dissociation of the protein [281] and molecular crowding along DNA [389] or within the cytoplasm [294].

The BHW model and its extensions provide a plausible mechanism for facilitated diffusion that has some support from experimental studies, which demonstrate that proteins do indeed slide along DNA [230, 389, 684]. In particular, recent advances in single-molecule spectroscopy means that the motion of fluorescently labeled proteins along DNA chains can be quantified with high precision, although it should be noted that most of these studies have been performed in vitro. A quantitative comparison of the BHW model with experimental data leads to a number of discrepancies, however. For example, it is usually assumed that $D_1 \approx D_3$ in order to obtain a sufficient level of facilitation. On the other hand, single-molecule measurements indicate that $D_1 \ll D_3$ [691]. Such experiments have also shown that $\tau_1 \gg \tau_3$, which is significantly different from the optimal condition $\tau_1 = \tau_3$. Hence the intermittent search process could actually result in a slowing down compared to pure 3D diffusion [281]. The BHW model also exhibits unphysical behavior in certain limits. These issues have motivated a number of alternative models of facilitated diffusion, as recently highlighted in [346].

1. *Electrostatic interactions*. One alternative hypothesis is that the observed fast association rates are due to electrostatic interactions between oppositely charged molecules and thus do not violate the 3D diffusion limit [249]. This is motivated by the theoretical result that the maximal association rate in Smoluchowski theory when there are long-range interactions between the reacting molecules is

$$k = 4\pi D a/\beta, \quad \beta = \int_a^\infty \mathrm{e}^{U(r)/k_B T} \frac{dr}{r^2},$$

where $U(r)$ is the interaction potential. The standard result is recovered when $U(r) = 0$ for $r > a$ [see Eq. (2.4.1)]. It follows that long-range attractive interactions can significantly increase diffusion-limited reaction rates. It has been further argued that in vitro experiments tend to be performed at low salt concentrations so that the effects of screening could be small. However, experimentally based estimates of the Debye length, which specifies the size of the region where electrostatic forces are important, indicate that it is comparable to the size of the target sequence. Hence, electrostatic forces are unlikely to account for facilitated diffusion.

2. *Colocalization.* Another proposed mechanism is based on the observation that in bacteria, genes responsible for producing specific proteins are located close to the binding sites of these proteins. This colocalization of proteins and binding sites could significantly speed up the search process by requiring only a small number of alternating 3D and 1D phases [440]. However, such a mechanism might not be effective in eukaryote cells, where transcription and translation tend to be spatially and temporally well separated. Moreover, colocalization breaks down in cases where proteins have multiple targets on DNA.
3. *Correlations*. Yet another theoretical mechanism involves taking into account correlations between 1D sliding and 3D bulk diffusion. These correlations reflect the fact that attractive interactions between a protein and nonspecific binding sites means that there is a tendency for a protein to return back to a neighborhood of the DNA site from which it recently dissociated [106, 702]. Although such interactions tend to slow down proteins moving along DNA, they also increase the local concentration of proteins absorbed to DNA. This suggests that facilitated diffusion can occur at intermediate levels of protein concentration and protein-DNA interactions.

7.6.2 Optimal Unbiased Search on a 1D Track

Consider a single-motor-driven particle moving along a one-dimensional track of length $2L$ (see Fig. 7.33). Suppose that at time $t = 0$ the particle is at some position y on the track. Within the interior of the track, $-L < x < L$, the particle can be in one of three states labeled by $n = 0, \pm$: stationary ($n = 0$), moving to the right (anterograde) with speed v_+ ($n = +$), or moving to the left (retrograde) with speed v_- ($n = -$). Transitions between the three states are governed by a discrete Markov process. We further assume that there is a hidden target of width $2a$ at the center of the domain. If the particle is within a distance a of the target and is in the stationary state, then the particle can detect or, equivalently, be absorbed by the target at a rate k. We assume throughout that $a \ll L$. Let $Z(t)$ and $N(t)$ denote the random position and state of the particle at time t and define $\mathbb{P}(x,t,n \mid y,0,m)dx$ as the joint probability that $x \leq Z(t) < x + dx$ and $N(t) = n$ given that initially the particle was at position $Z(0) = y$ and was in state $N(0) = m$. Setting $p_n(x,t) \equiv \mathbb{P}(x,t,n|y,0,+)$

with initial condition $p_n(x,0) = \delta(x-y)\delta_{n,+}$, we have a slightly modified version of the three-state model (7.4.2):

$$\partial_t p_+(x,t) = -v\partial_x p_+(x,t) - \beta p_+(x,t) + \alpha p_0(x,t), \tag{7.6.14a}$$
$$\partial_t p_-(x,t) = v\partial_x p_-(x,t) - \beta p_-(x,t) + \alpha p_0(x,t), \tag{7.6.14b}$$
$$\partial_t p_0(x,t) = \beta p_+(x,t) + \beta p_-(x,t) - 2\alpha p_0(x,t) - k\chi(x)p_0(x,t). \tag{7.6.14c}$$

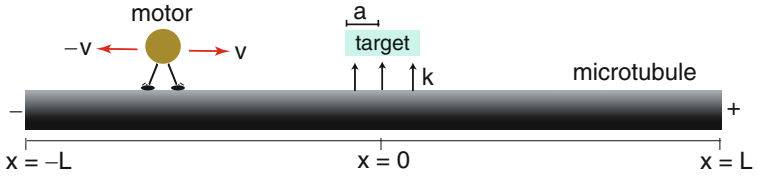


Fig. 7.33: Schematic diagram illustrating a model of a motor-driven particle moving along a one-dimensional track of length L. The particle can transition between two motile states with speeds $\pm v$ and a stationary state. A hidden target of half-width a is located at the center of the domain. The particle can be absorbed by the target at a rate k when in the stationary state

Here α, β are the transition rates between the stationary and mobile states and χ is the target indicator function

$$\chi(x) = \begin{cases} 1, & \text{if } |x| < a \\ 0, & \text{otherwise.} \end{cases} \tag{7.6.15}$$

Note that the motion is unbiased. Equations (7.6.14) are supplemented by a reflecting boundary conditions at each end:

$$p_-(x,t) = p_+(x,t), \quad x = \pm L. \tag{7.6.16}$$

The efficacy of the search process can be characterized in terms of the MFPT to find (be absorbed by) the target. There are two alternative methods for calculating the MFPT, one based on Laplace transforming the forward CK equation (7.6.14) and the other based on solving the corresponding backward equation. We will describe the latter approach here as developed by Loverdo et al. [398]. The backward CK equation is given by

$$\partial_t Q_+ = v\partial_y Q_+ - \beta[Q_+ - Q_0], \tag{7.6.17a}$$
$$\partial_t Q_- = -v\partial_y Q_- - \beta[Q_- - Q_0], \tag{7.6.17b}$$
$$\partial_t Q_0 = \alpha[Q_+ + Q_- - 2Q_0] - k\chi(y-X)Q_0, \tag{7.6.17c}$$

where $Q_m(y,t) = \mathbb{P}(x,t,0|y,0,m)$. Let $S_m(y,t)$ be the probability that the particle is absorbed by the target after time t given that it started at y in state m. That is,

$$S_m(y,t) = k\int_t^\infty \int_a^a \mathbb{P}(x,t',0|y,0,m)dxdt'. \tag{7.6.18}$$

Integrating Eq. (7.6.17) with respect to x and t and using

$$\partial_t S_m(y,t) = -k\int_{-a}^a \mathbb{P}(x,t,0|y,0,m)dx,$$

we find that

$$\partial_t S_+ = v\partial_y S_+ + \beta(S_0 - S_+), \tag{7.6.19a}$$
$$\partial_t S_- = -v\partial_y S_- + \beta(S_0 - S_-), \tag{7.6.19b}$$
$$\partial_t S_0 = \alpha[S_+ + S_- - 2S_0] - k\chi(y-X)S_0. \tag{7.6.19c}$$

Let $T_m(y)$ be the MFPT to find the target given that the particle is at $x = y$ and state m at $t = 0$. Then

$$T_m(y) = -\int_0^\infty t\partial_t S_m(y,t)dt = \int_0^\infty S_m(y,t)\mathrm{d}t \tag{7.6.20}$$

after integration by parts. It follows that T_m evolves according to the equations

$$v\partial_y T_+ + \beta(T_0 - T_+) = -1, \tag{7.6.21a}$$
$$-v\partial_y T_- + \beta(T_0 - T_-) = -1, \tag{7.6.21b}$$
$$\alpha(T_+ + T_-) - (2\alpha + k\chi(y-X))T_0 = -1. \tag{7.6.21c}$$

Solving Eq. (7.6.21c) for T_0 yields

$$T_0(y) = u(y)(\alpha[T_+(y) + T_-(y)] + 1), \tag{7.6.22}$$

where

$$u(y) = \frac{1}{2\alpha + k\chi(y-X)}. \tag{7.6.23}$$

Substituting Eq. (7.6.22) into Eq. (7.6.21a,b) gives

$$\partial_y T_+(y) + \frac{\beta}{v}[(\alpha u(y) - 1)T_+(y) + \alpha u(y)T_-(y)] = -\frac{\beta}{v}\left(\frac{1}{\beta} + u(y)\right), \tag{7.6.24a}$$
$$\partial_y T_-(y) - \frac{\beta}{v}[\alpha u(y)T_+(y) + (\alpha u(y) - 1)T_-(y)] = \frac{\beta}{v}\left(\frac{1}{\beta} + u(y)\right). \tag{7.6.24b}$$

It is now necessary to solve for $T_\pm(y)$ in the three regions: $-L < y < -a$, $-a < y < a$, and $a < y < L$. The solution in each of these regions will have two unknown

integration constants so that we require six conditions. Two are given by the boundary conditions

$$T_+(y) = T_-(y), \quad y = \pm L, \tag{7.6.25}$$

whereas the other four are obtained by requiring continuity in $T_+(y)$ and $T_-(y)$ at $y = \pm a$ (see Ex. 7.11). Suppose that the particle starts at a random position inside the domain and is initially in the $+$ state. We then define the averaged MFPT τ_1 according to

$$\tau_1 = \frac{1}{2L}\int_{-L}^{L} T_+(y)dy.$$

From the analysis of $T_\pm(y)$, one finds that [29] (see Ex. 7.11)

$$\tau_1 = \frac{1}{L}\left(\frac{1}{\beta}+\frac{1}{2\alpha}\right)\left(\left[\frac{\beta}{v}\right]^2\frac{2(L-a)^3}{3}+\frac{\beta}{v}\sqrt{\frac{2\alpha+k}{k}}(L-a)^2\coth(\Lambda a)\right) + \frac{2\alpha+k}{k}\left(\frac{1}{\beta}+\frac{1}{2\alpha+k}\right)\frac{L-a}{L}, \tag{7.6.26}$$

where

$$\Lambda = \frac{\beta}{v}\sqrt{\frac{k}{2\alpha+k}}.$$

This expression can be simplified by taking $L \gg a$ and $a\beta/v \ll 1$:

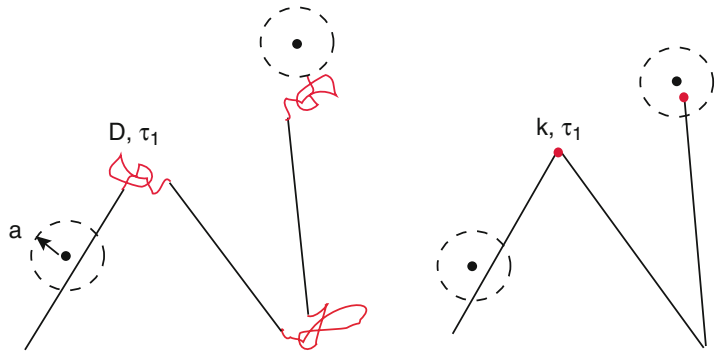


Fig. 7.34: Two models of target detection: the particle alternates between slow reactive phases of mean duration τ_1 and fast nonreactive ballistic phases of mean duration τ_2. *Left*: the slow reactive phase is diffusive and detection is infinitely efficient. *Right*: the slow reactive phase is static and detection takes place with finite rate k

$$\tau_1 \approx \left(\frac{1}{\beta}+\frac{1}{2\alpha}\right)\left(\left[\frac{\beta}{v}\right]^2\frac{2L^2}{3}+\frac{2\alpha+k}{k}\frac{L}{a}\right). \tag{7.6.27}$$

It is then straightforward to show that τ_1 has a global minimum as a function of the parameters α, β, which occurs at the values (α^*, β^*) with

$$\frac{1}{\alpha^*} = 2\sqrt{\frac{a}{vk}}\left(\frac{L}{12a}\right)^{1/4}, \quad \frac{1}{\beta^*} = \frac{a}{v}\sqrt{\frac{L}{3a}}.$$

It is possible to extend the above analysis to a modified target detection scheme, in which the particle slowly diffuses in the search phase and is immediately absorbed by the target if it enters the target domain ($k \to \infty$) (see Fig. 7.34). In this case, an intermittent search scheme is more efficient than pure diffusion provided that $D/v \ll a$, where D is the diffusivity of the particle in the search phase, v is its speed in a ballistic non-search phase, and a is the size of the target [398]. Interestingly the optimal time spent in the non-search phase, $1/\beta^*$, is independent of the particular target detection mechanism. The existence of a minimum search time for unbiased intermittent search in a bounded domain also extends to higher spatial dimensions [31, 70, 398] and to more detailed molecular motor models such as ToW [472, 474]. However, in these more complicated cases, the calculation of the MFPT becomes considerably more difficult unless some approximation scheme is used such as the QSS reduction outlined in Sect. 7.4 [471, 473] (see below).

7.6.3 Biased Intermittent Search

The existence of an optimal search strategy breaks down if one considers a biased search process and allow for the possibility of failure to find the target [68, 471]. In order to illustrate this, we modify the previous 1D model (7.6.14) along the following lines. First the transition rates from the left- and right-moving states are taken to be different, $\beta \to \beta_\pm$ with $\beta_+ < \beta_-$. The stochastic process is then biased in the anterograde direction since the particle tends to spend more time in the right-moving state. (One could also take the velocities in the two directions to be different.) Unidirectional transport is obtained in the limit $\beta_- \to \infty$. Second, the possibility of failure is incorporated into the model by considering a 1D track of length L with a reflecting boundary at $x = 0$ and an absorbing boundary at $x = L$:

$$p_-(0,t) = p_+(0,t), \quad p_-(L,t) = 0. \tag{7.6.28}$$

The absorbing boundary takes into account the fact that a motor particle can be degraded or absorbed by other targets downstream to the given target. Third, the particle always starts from the end $x = 0$ (which could be close to the cell nucleus) and the target is at some unknown location X with $0 < X - a < X + a < L$. In contrast to the unbiased case, the efficacy of the search process is characterized by two quantities. Let $J(t)$ denote the probability flux due to absorption by the target at X:

$$J(t) = k\int_{X-a}^{X+a} p_0(x,t)dx. \tag{7.6.29}$$

Define the hitting probability Π to be the probability that the particle eventually finds the target, that is, it is absorbed somewhere in the interval $X-a \leq x \leq X+a$ rather than at the end $x=L$:

$$\Pi = \int_0^\infty J(t)dt. \tag{7.6.30}$$

The conditional MFPT [523] T is then defined to be the mean time it takes for the particle to find the target given that it is not absorbed at $x=L$:

$$T = \frac{\int_0^\infty tJ(t)dt}{\int_0^\infty J(t)dt}. \tag{7.6.31}$$

The two quantities T and Π determine the efficiency of the stochastic search process.

Clearly it would be advantageous for the particle to minimize the search time T and maximize the hitting probability Π. However, these two requirements compete with each other so that, in contrast to unbiased intermittent search with $\Pi=1$, there is not a single optimal search strategy. This can be seen heuristically in the case of unidirectional transport where the particle is either stationary or undergoes anterograde motion. Here the particle can reach the target more quickly by having a higher probability of being in the mobile state. However, this also increases the chance of overshooting the target without detecting it, thus reducing the hitting probability. It could be argued that the only important factor is minimizing the MFPT irrespective of the hitting probability, since active transport typically involves multiple motor–cargo complexes. However, a low hitting probability would require more resources, which costs the cell energy. In the case of unidirectional transport, Eq. (7.6.14) reduce to

$$\partial_t p_+ = -v\partial_x p_+ + \alpha p_0 - \beta p_+, \tag{7.6.32a}$$

$$\partial_t p_0 = \beta p_+ - \alpha p_0 - k\chi(x-X)p_0. \tag{7.6.32b}$$

Note that there is no need to introduce any supplementary boundary conditions, since the particle cannot return to the origin nor find the target once it has crossed the point $x=X+a<L$. The corresponding backward equations are

$$\partial_t Q_+ = v\partial_y Q_+ - \beta[Q_+ - Q_0], \tag{7.6.33a}$$

$$\partial_t Q_0 = \alpha[Q_+ - Q_0] - k\chi(y-X)Q_0, \tag{7.6.33b}$$

where $Q_m(y,t) = \mathbb{P}(x,t,0|y,0,m)$.

Let $\gamma_m(y,t)$ be the total probability that the particle is absorbed by the target after time t given that it started at y in state m. That is,

$$\gamma_m(y,t) = k\int_t^\infty \int_{X-a}^{X+a} \mathbb{P}(x,t',0|y,0,m)dxdt'. \tag{7.6.34}$$

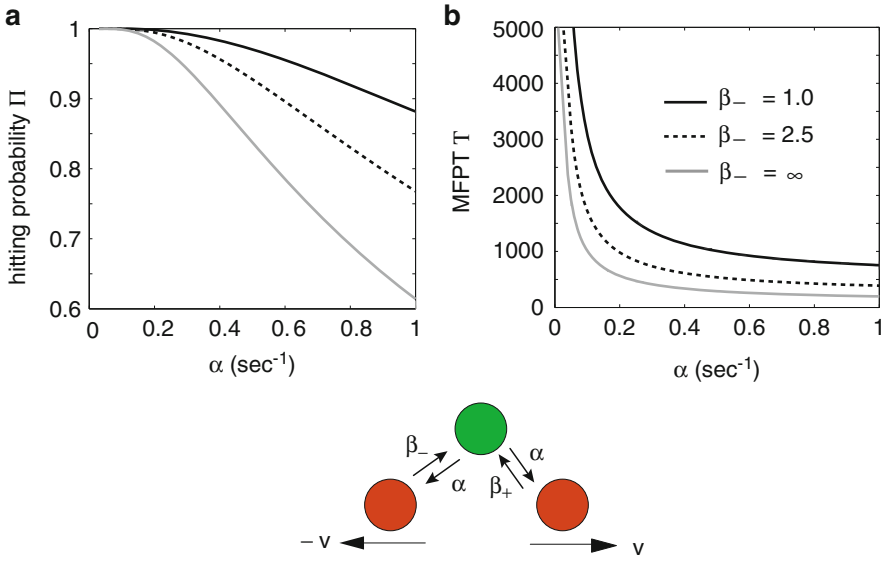


Fig. 7.35: Partially biased anterograde transport. (**a**) The hitting probability Π and (**b**) the MFPT T are plotted as functions of the parameter α for fixed $\beta_+ = 1\,\mathrm{s}^{-1}$ and various values of β_-: *solid black curve* ($\beta_- = 1.5\,\mathrm{s}^{-1}$), *dashed curve* ($\beta_- = 2.5\,\mathrm{s}^{-1}$), and *solid gray curve* (unidirectional). Other parameter values are $X = 10\,\mu\mathrm{m}$, $L = 20\,\mu\mathrm{m}$, $a = 1\,\mu\mathrm{m}$, $k = 0.05\,\mathrm{s}^{-1}$, and $v_\pm = 0.1\,\mu\mathrm{m}\,\mathrm{s}^{-1}$. These values are extracted from experimental studies of mRNA transport [153, 342, 544]

Integrating Eqs. (7.6.33a) and (7.6.33b) with respect to x and t and using $\partial_t \gamma_m(y,t) = -k\int_{X-a}^{X+a} \mathbb{P}(x,t,0|y,0,m)dx$, we find that

$$\partial_t \gamma_+ = v\partial_y \gamma_+ + \beta(\gamma_0 - \gamma_+), \tag{7.6.35a}$$

$$\partial_t \gamma_0 = \alpha[\gamma_+ - \gamma_0] - k\chi(y-X)\gamma_0. \tag{7.6.35b}$$

The probability $\gamma_m(y,t)$ can be used to define the hitting probability

$$\pi_m(y) = \gamma_m(y,0) \tag{7.6.36}$$

and the conditional MFPT $T_m(y)$,

$$T_m(y) = -\int_0^\infty t\frac{\partial_t \gamma_m(y,t)}{\gamma_m(y,0)}\mathrm{d}t = \frac{\int_0^\infty \gamma_m(y,t)\mathrm{d}t}{\gamma_m(y,0)} \tag{7.6.37}$$

after integration by parts. It follows from the definitions (7.6.30) and (7.6.31) that

$$\Pi = \pi_+(0), \quad T = T_+(0). \tag{7.6.38}$$

Proceeding along analogous lines to the unbiased case, one finds that [68]

$$\Pi = 1 - \mathrm{e}^{-2\lambda a}, \quad T = (X-a)\mu_1 - \frac{2a\mu_2}{\mathrm{e}^{2\lambda a}-1} + \mu_3,$$

where

$$\mu_1 = \frac{1}{v}\left(1+\frac{\beta}{\alpha}\right), \quad \mu_2 = \frac{1}{v}\left(1+\frac{\alpha\beta}{(\alpha+k)^2}\right), \quad \mu_3 = \frac{\alpha+\beta+k}{\beta k}, \quad \lambda = \frac{\beta}{v}\frac{k}{\alpha+k}.$$

The calculation of Π and T can also be carried out for the full three-state model, although the analysis is considerably more involved [68]. The results are illustrated in Fig. 7.35. It can be seen that increasing the parameter α, which controls how much time the particle spends in the stationary search mode, decreases both the hitting probability and the conditional MFPT. Similarly, increasing the parameter β_+, which controls how much time the particle spends in the anterograde mobile state, increases both the hitting probability and the MFPT. During unidirectional motion average velocities are found in the range 0.05–0.2 μm s^{-1}, whereas the duration of a moving phase tends to be in the range 1–10 s. Dendrites in cultured cells range in length from 10 to 100 μm.

7.6.4 Effects of Local Chemical Signaling

Let us now consider a much more general model of motor-driven search by incorporating a hidden target into the PDE model (7.4.4)

$$\frac{\partial p}{\partial t} = -v_n \frac{\partial [p(x,n,t)]}{\partial x} + \sum_{n'=1}^{N} A(n,n';x)p(x,n',t) - k_n \chi(x-X), \quad (7.6.39)$$

where k_n is the rate of target absorption in the nth internal state. Thus the flux into the target is

$$J(t) = \sum_{n=1}^{N} k_n \int_{X-a}^{X+a} p(x,n,t)dx. \quad (7.6.40)$$

In general k_n will only be nonzero for a subset of states. For example, in the three-state model the unbound stationary or diffusing state is identified as the search state. In the case of the more biophysically realistic ToW model (Sect. 4.4), the identification of the search states is more complicated. The simplest scenario is that the cargo locates its target after it becomes fully detached from the microtubule and diffuses within distance of its target, where it binds to scaffolding proteins and is separated from its molecular motors. However, if many molecular motors are bound to the cargo, the waiting time between diffusive searching events can be too large to reliably deliver the cargo. Moreover, if the cargo is large so that its diffusivity is low or the cargo is moving through a crowded and confined domain, diffusive motion

may be restricted, preventing the cargo from reaching the target. Another possibility is that subcellular machinery is present to detach the cargo from its motors or inhibit the activity of the motors so that scaffolding proteins can bind to and sequester the cargo. Delivery then changes from a diffusion-limited reaction to a waiting time that depends on a reaction occurring between the motor–cargo complex and biomolecules (either freely diffusing or anchored) local to the target while the motor-cargo complex is moving along the microtubule. If details of the localization mechanism are unknown, then the simplest model is to assume that this waiting time is approximately exponential and to associate a target detection rate k_n with each motor state. The model can be simplified further by assuming that detection is unlikely while only one species of motors is engaged and pulling the cargo at its maximum (forward or backward) velocity. This suggests assigning a single target detection rate k to those states that have sufficiently low speeds [474]. Thus, $k_{(n_+,n_-)} = k\Theta(v_h - v(n_+,n_-))$, where $v(n_+,n_-)$ denotes the velocity when n_+ kinesin and n_- dynein motors are attached to the track and v_h is a velocity threshold.

It is straightforward to extend the QSS approximation of Sect. 7.4 in the presence of a target, and one finds that [473]

$$\frac{\partial C}{\partial t} = -\frac{\partial}{\partial x}(VC) + \frac{\partial}{\partial x}\left(D\frac{\partial C}{\partial x}\right) - \lambda\chi(x-X)C, \tag{7.6.41}$$

with the drift V and diffusion coefficient D given by Eq. (7.4.11) and the effective detection rate is

$$\lambda = \sum_{n=1}^{N} k_n \rho(x,n). \tag{7.6.42}$$

There are now three effective parameters that describe the random search process: the drift V, the diffusivity D, and the target detection rate λ. Each of these parameters are themselves functions of the various cargo velocities, transition rates, and target detection rates contained in the full model. The hitting probability and MFPT are still given by Eqs. (7.6.30) and (7.6.31) except that now the flux is

$$J(t) = \lambda \int_{X-a}^{X+a} C(x,t)dx. \tag{7.6.43}$$

In general, one finds that there is a play-off between minimizing the MFPT and maximizing the hitting probability [68, 471, 473]. One way to enhance the efficiency of the search process is for the target to generate a local chemical signal that increases the probability of finding the target without a significant increase in the MFPT. This issue has been explored by incorporating a local ATP or tau signal into the ToW model of Sect. 4.4 and carrying out a QSS reduction along the lines outlined above [472, 474]. The possible role of ATP is based on the observation that the stall force and other single-motor parameters are strongly dependent on the level of [ATP]. Since ATP concentration ([ATP]) is heavily buffered, a small region of intense ATP phosphorylation around a target could create a sharp, localized [ATP] gradient, which would significantly slow down a nearby motor complex, thus increasing the

chances of target detection. Here we consider a mechanism based on local tau signaling [472], based on the ToW model with the tau concentration-dependent kinesin

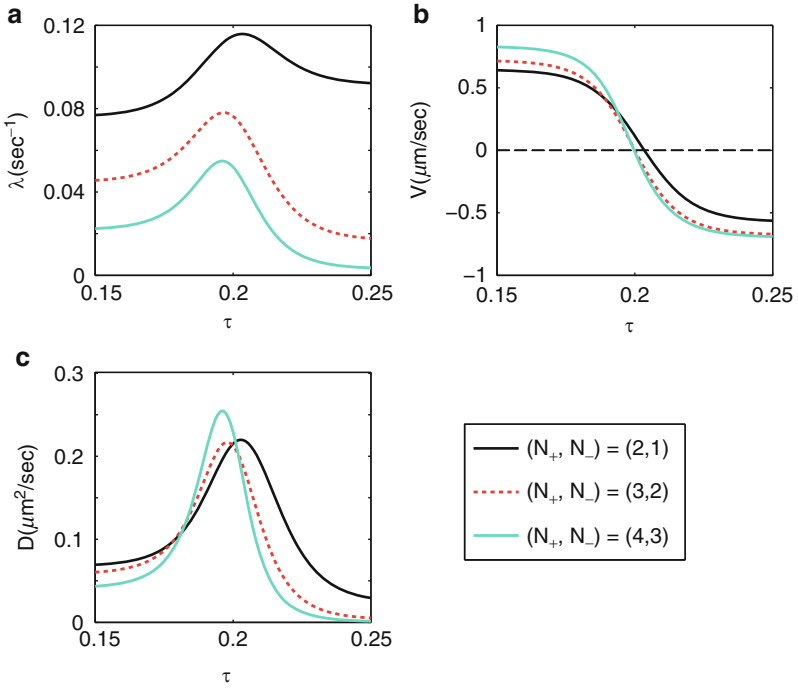


Fig. 7.36: Effects of tau concentration on the tug-of-war model with N_+ kinesin motors and N_- dynein motors. The stall force F_S, forward velocity v_f, and unbinding rate $\bar{\gamma}$ are given by Eqs. (4.4.14)–(4.4.16) with $[\mathrm{ATP}] = 10^3\,\mu\mathrm{M}$. The other single-motor parameters are [457] $F_d = 3\,\mathrm{pN}$, $\gamma_0 = 1\,\mathrm{s}^{-1}$, $\bar{\pi} = 5\,\mathrm{s}^{-1}$, and $v_b = 0.006\,\mu\mathrm{m\,s}^{-1}$. The corresponding parameters of the FP equation are obtained using a QSS reduction and plotted as a function of τ. (**a**) Effective capture rate λ. (**b**) Drift velocity V. (**c**) Diffusivity D

binding rate Eq. (4.4.17). Carrying out the QSS reduction of the tug-of-war model then leads to the FP equation (7.6.41) with τ-dependent drift V, diffusivity D, and capture rate λ as illustrated in Fig. 7.36. The most significant alteration in the behavior of the motor complex is the change in the drift velocity V as a function of τ. The drift velocity switches sign when τ is increased past a critical point. That is, by reducing the binding rate of kinesin, the dynein motors become dominant, causing the motor complex to move in the opposite direction. The effects of local changes in τ concentration on the efficiency of random search can now be determined by assuming that within range of the target, $|x-X| < a$, $\tau = \tau_1 > \tau_0$, whereas $\tau = \tau_0$ outside the target, $|x-X| > a$. Carrying out the QSS reduction of the tug-of-war model then leads to the FP equation (7.4.10) with x-dependent drift and diffusivity:

$$V(x) = V_0 + \Delta V \chi(x), \quad D(x) = D_0 + \Delta D \chi(x), \tag{7.6.44}$$

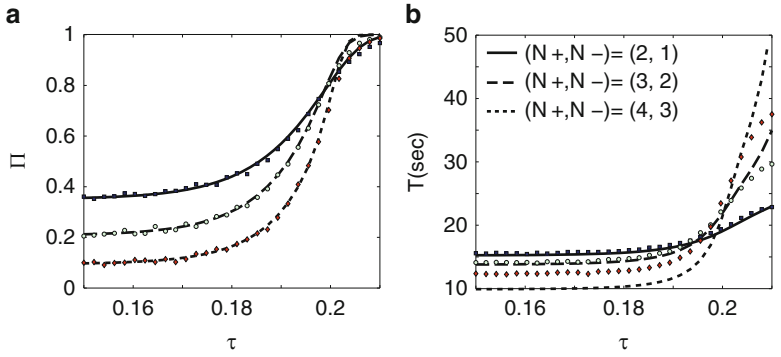


Fig. 7.37: Effect of adding tau to the target on the capture probability Π and MFPT T using parameters from Fig. 7.36. (**a**) The analytical approximation P (*solid line*) and results from Monte Carlo simulation. (**b**) The analytical approximation T along with averaged Monte Carlo simulations. The synaptic trap is located at $X = 10\,\mu\text{m}$, the trapping region has radius $a = 2\,\mu\text{m}$, and the MT track has length $L = 20\,\mu\text{m}$. The capture rate is taken to be $k_0 = 0.5\,\text{s}^{-1}$

where $\chi(x)$ is the indicator function defined in Eq. (7.6.15), $V_0 = V(\tau_0), D_0 = D(\tau_0)$, $\Delta V = V(\tau_1) - V_0$, and $\Delta D = D(\tau_1) - D_0$. Solving the piecewise continuous FP equation then determines the hitting probability Π and MFPT T as functions of τ_1 for fixed τ_0. In Fig. 7.37, the hitting probability Π and the MFPT T are plotted as a function of τ_1. As τ_1 is increased above the critical level $\tau_0 = 0.19$, there is a sharp increase in Π but a relatively small increase in the MFPT, confirming that τ can improve the efficacy of the search process.

One interesting effect of a local increase in MAPs is that it can generate stochastic oscillations in the motion of the motor complex [474], see Fig. 7.38. As a kinesin-driven cargo encounters the MAP-coated trapping region, the motors unbind at their usual rate and can't rebind. Once the dynein motors are strong enough to pull the remaining kinesin motors off the microtubule, the motor complex quickly transitions to $(-)$ end-directed transport. After the dynein-driven cargo leaves the MAP-coated region, kinesin motors can then reestablish $(+)$ end-directed transport until the motor complex returns to the MAP-coated region. This process repeats until the motor complex is able to move forward past the MAP-coated region. Interestingly, particle tracking experiments have observed oscillatory behavior during mRNA transport in dendrites [153, 544]. In these experiments, motor-driven mRNA granules move rapidly until encountering a fixed location along the dendrite where they slightly overshoot then stop, move backward, and begin to randomly oscillate back and forth. After a period of time, lasting on the order of minutes, the motor-driven mRNA stops oscillating and resumes fast ballistic motion. Calculating the mean time to escape the target can be formulated as a FPT problem, in which the particle starts at $x = x_0$ and has to make a rare transition to the unstable fixed point at $x = x_*$. As in the analogous problem of stochastic action potential generation the QSS diffusion approximation breaks down for small ε, and one has to use singular perturbation methods. The details in the case of the three-state model can be found elsewhere [475].

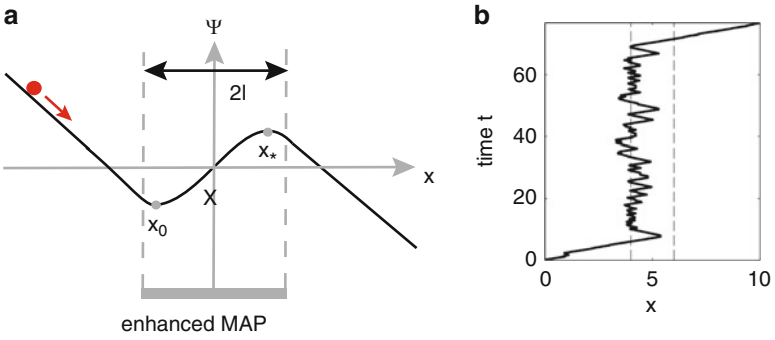


Fig. 7.38: Diagram showing (**a**) the effective potential well created by a region of tau coating an MT and (**b**) a representative trajectory showing random oscillations within the well

7.6.5 Virus Trafficking

An interesting example of random intermittent search in 2D or 3D is given by virus trafficking. An animal virus typically invades a mammalian cell by first undergoing membrane endocytosis from the exterior to the interior of the cell. It then has to navigate the crowded cytoplasm without being degraded in order to reach a nuclear pore and deliver its DNA to the cell nucleus [133]. SPT has established that virus trajectories within the cytoplasm consist of a succession of free or confined diffusion and ballistic periods involving active transport along microtubules or actin networks [60]. A macroscopic computational model of the trafficking of a population of viruses has been developed based on the law of mass action, which takes into account cell geometry but neglects stochastic effects [142]. More recently, in a series of papers, Holcman and collaborators [266, 368, 369] have developed a stochastic model of a single virus trafficking inside a cell, which involves reducing an intermittent search model to an effective Langevin equation, and using the latter to calculate the mean time to reach a nuclear pore based on a narrow escape problem (see Sect. 7.2). The basic structure of a 2D version of the latter model is shown in Fig. 7.39.

Following [368], the cell is treated as a radially symmetric disk consisting of an annular region of cytoplasm of outer radius R and inner radius δ, surrounding a central nuclear disk. N microtubules radiate outward from the nucleus to the cell membrane and are assumed to be distributed uniformly so that the angle between two neighboring microtubules is $\Theta = 2\pi/N$. (A two-dimensional description of a cell would be reasonable in the case of cultured cells that are flattened due to adhesion to the substrate.) The motion of a virus particle alternates between diffusive motion within a wedge region $\hat{\Omega}$ subtending an angle Θ at the origin and binding to one of the two microtubules at the boundary of the wedge. Suppose that a virus particle starts at some radius $r_0 < R$ and arbitrary angle within such a wedge. Let $\tau(r_0)$ denote the MFPT for the particle to bind to a microtubule, and let $\rho(r_0)$

be the mean radial position on the microtubule. Suppose that the particle moves with a fixed speed v for a time T towards the nucleus before being released to a new position with radius r_1 and arbitrary angle within another wedge. It follows that $r_1 = \rho(r_0) - vT$. Treating the domain $\hat{\Omega}$ as an open wedge by ignoring the reflecting boundary at $r = R$, it can be shown that if $\Theta \ll 1$, then [369]

$$\tau(r_0) \approx r_0^2\Theta^2/12D, \quad \rho(r_0) \approx r_0(1+\Theta^2/12).$$

The reduction method of [368, 369] is to assume that on a coarse-grained time scale the random intermittent motion of the virus can be approximated by a Langevin equation with a radial drift vector:

$$\frac{d\mathbf{r}}{dt} = b(r)\frac{\mathbf{r}}{|\mathbf{r}|} + \sqrt{2D}d\xi dt. \tag{7.6.45}$$

In order to estimate the drift function $b(r)$, the MFPT $\hat{\tau}(r_0)$ for the effective Langevin particle to start at r_0 and end at r_1 is calculated using the standard theory of FPTs (see Sect. 2.3) and then compared to $\tau(r_0)$. First, $\hat{\tau}(r_0)$ satisfies the equation

$$D\nabla^2\hat{\tau} - b(r)\nabla\hat{\tau} = -1,$$

with boundary conditions

$$\frac{d\hat{\tau}}{dr}(R) = 0, \quad \hat{\tau}(r_1) = 0.$$

As a further simplification, it is assumed that $b(r)$ varies slowly with r so that $b(r) \approx b(r_0)$, leading to the solution

$$\hat{\tau}(r_0) = \int_{r_1}^{r_0}\int_v^R \frac{ue^{-b(r_0)[u-v]/D}}{Dv}dudv.$$

Assuming that $D \ll 1$ the Laplace method can be used to evaluate the integral with respect to u, giving $\hat{\tau}(r_0) \approx (r_0 - r_1)/b(r_0)$. Finally, setting $\hat{\tau}(r_0) = \tau(r_0) + T$ yields

$$b(r_0) = \frac{r_0 - r_1}{\tau(r_0)+T} = \frac{d - r_0\Theta^2/12}{T + r_0^2\Theta^2/12D}. \tag{7.6.46}$$

A more detailed calculation of the effective drift function $b(r)$ under less restrictive assumptions can be found in [369].

Having reduced the effective motion of the virus to a Langevin equation, the probability that the virus arrives at a nuclear pore before being degraded at a rate k_0 can now be calculated by solving a narrow escape problem. The associated FP equation takes the form

$$\frac{\partial p}{\partial t} = D\nabla^2 p(\mathbf{r},t) - \nabla\cdot\mathbf{b}(\mathbf{r})p(\mathbf{r},t) - k_0 p(\mathbf{r},t) \tag{7.6.47}$$

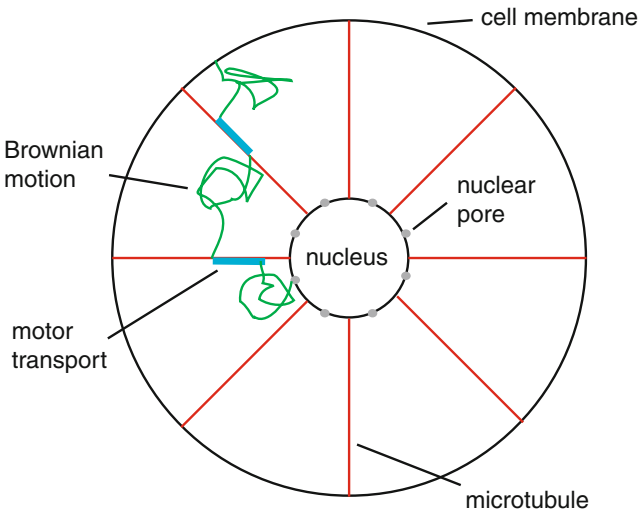


Fig. 7.39: Model of Lagache et al. [368]. Diagram of a 2D radially symmetric cell with radially equidistant microtubles. A virus trajectory is shown that alternates between ballistic motion along a microtubule and diffusion the cytoplasm. Trajectory starts at the cell membrane and ends at a nuclear pore

on the annular region Ω of Fig. 7.39, together with the boundary conditions

$$p(\mathbf{r},t) = 0, \mathbf{r} \in \partial N_a, \quad J(\mathbf{r},t) \cdot \mathbf{n} = 0, \mathbf{r} \in \partial\Omega - \partial N_a.$$

The boundary $\partial\Omega$ of the annulus is taken to be reflecting everywhere except for the surface ∂N_a of the nucleus occupied by nuclear pores, which are taken to be perfect absorbers. Asymptotic analysis then shows that the hitting probability $\mathscr{P}$ and conditional MFPT $\mathscr{T}$ are [368, 369]

$$\mathscr{P} = \frac{b(\delta)}{b(\delta) + 2\delta k_0 \nu}, \quad \mathscr{T} = \frac{2\delta\nu}{2\delta k_0 \nu + b(\delta)}, \tag{7.6.48}$$

where $\nu = \log(1/\varepsilon)$ with ε the fraction of the nucleus covered by nuclear pores.

7.7 Exercises

Problem 7.1 (Diffusion–trapping model with receptor binding). Consider an extension of the 1D diffusion–trapping model given by Eqs. (7.1.2a) and (7.1.2b), which includes the effects of receptor binding to scaffolding proteins within the

spine. That is, let R_i and B_i denote, respectively, the probability of an unbound and a bound receptor in the ith spine. Replace Eq. (7.1.2b) by the pair of equations

$$\frac{dR_j}{dt} = lh[p_j - R_j/A] + \alpha B_j - \beta R_j, \quad \frac{dB_j}{dt} = -\alpha B_j + \beta R_j,$$

where α, β are the rates of unbinding and binding. Calculate the MFPT to travel a certain distance from one end and use this to obtain the following expression for the effective diffusivity:

$$D_{\text{eff}} = D\left[1 + \frac{A}{ld}\left(1 + \frac{\beta}{\alpha}\right)\right]^{-1}.$$

Describe what happens when $\beta \gg \alpha$.

Problem 7.2 (Steady-state analysis of diffusion–trapping model). Consider a slight modification of the diffusion–trapping model in a dendrite of length L and circumference l, in which receptors within a spine degrade at a rate γ:

$$\frac{\partial p}{\partial t} = D\frac{\partial^2 p}{\partial x^2} - \sum_{j=1}^{N} h[p_j - R_j/A]\delta(x - x_j),$$
$$\frac{dR_j}{dt} = lh[p_j - R_j/A] - \gamma R_j$$

Moreover, suppose that newly synthesized receptors enter the end $x = 0$ (close to the soma) at a rate σ. The boundary conditions become

$$D\left.\frac{\partial p}{\partial x}\right|_{x=0} = -\sigma, \quad D\left.\frac{\partial p}{\partial x}\right|_{x=L} = 0$$

(a) Show that the steady-state receptor distribution satisfies

$$0 = D\frac{\partial^2 p}{\partial x^2} - \beta\sum_{j=1}^{N} p_j\delta(x - x_j)$$

and determine the constant β. Using the boundary conditions, derive the conservation condition

$$\sigma = \beta\sum_{j=1}^{N} p_j.$$

(b) Introducing the Neumann Green's function

$$\frac{d^2G(x,y)}{dy^2} = -\delta(x - y) + L^{-1},$$

with reflecting boundary conditions at the ends $x = 0, L$, derive the following implicit equation:

$$p(x) = \chi - \beta \sum_j G(x,x_j) p_j + \frac{\sigma}{D} G(x,0).$$

The constant χ is determined from the conservation condition $\sigma = \beta \sum_{j=1}^N p_j$.

(c) By setting $x = x_i$ in the result of part (b), show that

$$p_i = \sum_j M_{ij}(\chi + \sigma G(x_j,0)/D), \quad \mathbf{M}_{ij}^{-1} = G(x_i,x_j) + \delta_{ij}.$$

Use this to determine χ.

(d) Consider a set of N identical spines with uniform spacing $d = L/N$ such that $x_j = jd$, $j = 1,\ldots,N$. Using the results of parts (b) and (c), plot $p(x)$ as a function of x for the three cases $D = 0.5, 0.1, 0.05\,\mu\text{m}^2\,\text{s}^{-1}$, with $L = 200\,\mu\text{m}$, $l = 1\,\mu\text{m}$, $N = 200$, spine spacing $d = 1\,\mu\text{m}$, $\sigma = 1\,\text{s}^{-1}$, $h = 10^{-3}\,\mu\text{s}^{-1}$, and $\gamma = 10^{-4}\,\text{s}^{-1}$. Use the fact that

$$G(x,x') = \frac{L}{12}\left[h([x+x']/L) + h(|x-x'|/L)\right],$$

where $h(x) = 3x^2 - 6|x| + 2$.

(e) Consider a set of N identical spines with uniform spacing $d = L/N$ such that $x_j = jd$, $j = 1,\ldots,N$, and rewrite the steady-state equation in the form

$$0 = D\frac{d^2p}{dx^2} - \beta\rho(x)p \tag{7.7.49}$$

where $\rho(x) = \sum_{j=1}^N \delta(x - jd)$. In the large N limit, take $\rho = 1/d$ and solve the resulting continuum equation. Compare with the results of part (d).

Problem 7.3 (Exact solutions of Fick–Jacobs equation). Consider the Fick–Jacobs equation (with $D_0 = 1$)

$$\frac{\partial P(x,t)}{\partial t} = \frac{\partial}{\partial x}\sigma(x)\frac{\partial}{\partial x}\frac{P(x,t)}{\sigma(x)}.$$

(a) Let $f(x) = \frac{1}{2}\ln\sigma(x)$. Performing the change of variables

$$A(x,t) = \mathrm{e}^{-f(x)}P(x,t) = \frac{P(x,t)}{\sqrt{\sigma(x)}},$$

show that A satisfies an equation of the form

$$\frac{\partial A(x,t)}{\partial t} = \frac{\partial^2 A}{\partial x^2} - V(x)A(x,t),$$

with

$$V(x) = f'^2(x) + f''(x) = \frac{1}{2\sigma(x)}\frac{d^2\sigma(x)}{dx^2} - \frac{1}{4}\left(\frac{\sigma'(x)}{\sigma(x)}\right)^2.$$

(b) Determine the effective potential $V(x)$ in the two cases (i) $\sigma(x) = \pi(1+\lambda x)^2$ (conical cross section) and (ii) $\sigma(x) = \sin^2(\gamma x)$. Hence, for each case, express the solution $P(x,t)$ in terms of the initial data $P(x,0)$ and the fundamental solution of the diffusion equation.

Problem 7.4 (Multibarrier model of ion channel transport). Suppose that an ion passing through a channel encounters a series of energy barriers and wells as shown in Fig. 7.40. Ion channel movement within the channel can then be modeled in terms of a sequence of hops over the barriers from one site (minimum) to the next. For simplicity, in the absence of a voltage drop across the membrane ($\Delta V = 0$), the barriers are identical, symmetric, and spaced uniformly along the channel. A voltage drop can then be superimposed to bias the jumps in a particular direction. Let α and β denote the right and left hopping rates across each barrier. Assuming that the rates are given by a standard Arrhenius formula (Sect. 2.3), we can write

$$\alpha = \omega e^{-(E - q\Delta V/2n)/k_B T}, \quad \beta = \omega e^{-(E + q\Delta V/2n)/k_B T},$$

where ω is some prefactor that depends on the detailed shape of the channel potential energy around the minimum and q is the ion charge. Here E is the height of the energy barrier when $\Delta V = 0$ and the symmetric barriers are lowered by an

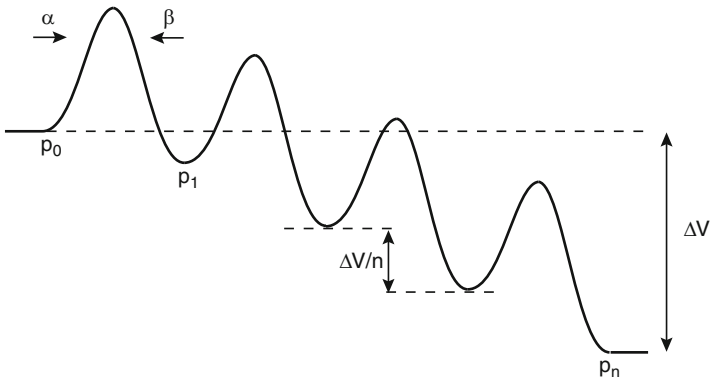


Fig. 7.40: Multibarrier potential of an ion channel

amount $\Delta V/2n$, where n is the number of barriers. Assume the system is in steady state so the flux J through the system is the same everywhere, and let p_i denote the probability that an ion is at the ith site.

(a) Iteratively solve the sequence of flux equations

$$J = \alpha p_i - \beta p_{i+1}, i = 0, \ldots, n-1,$$

with $p_0 = vc_a$ and $p_n = vc_b$. Here c_a and c_b are the bulk concentrations on the left and right sides of the channel, respectively, and v is related to a small volume element at the entry/exit of the channel. In particular, show that

$$J = v\omega \mathrm{e}^{-(E-q\Delta V/2n)/k_BT} \frac{(c_a - c_b\phi^n)(1-\phi)}{1-\phi^n}, \quad \phi = \mathrm{e}^{-q\Delta V/nk_BT}.$$

(b) Determine J when $c_a = c_b = c$. Plot the result flux as a function of voltage ΔV measured in units of $q/2k_BT$ for $n = 1, 2, 3$. Set the scale factor $cv\omega \mathrm{e}^{-E/k_BT} = 1$. What happens to the current-voltage curve for large n?

Problem 7.5 (Entropic gate model of nuclear pore transport). Consider the steady-state equation for the density $\rho(x)$ of cargo complexes in the NPC:

$$\frac{d}{dx}\left[\frac{d\rho}{dx} + \rho(x)\frac{dU(x)}{dx}\right] = 0,$$

supplemented by the boundary conditions $\rho(0) = \rho(L) = 0$. Here $U(x)$ is given by Fig. 7.16

(a) Solve the steady-state equation using the conditions $J(x) = -|J_0|$ for $0 < x < R$, $J(x) = \bar{J}$ for $R < x < L - R$ and $J(x) = J_L$ for $L - R < x < L$.

(b) Using $\bar{J} = J_L + J_{\mathrm{ran}}\rho(L-R)R$ and the solution in the region $L - R < x < L$ show that

$$D\rho(L-R) = \frac{R\bar{J}}{J_{\mathrm{ran}}R^2/D + 1}.$$

Now imposing continuity of $\rho(x)$ at $x = L - R$ and setting $\bar{J} = J_S - |J_0|$ derive the result

$$\frac{\bar{J}}{J_S} = \left[1 + \frac{1}{J_{\mathrm{ran}}R^2/D+1} + \frac{1}{R}\int_R^{L-R} \mathrm{e}^{-U(x')}dx'\right]^{-1}.$$

Problem 7.6 (Discrete model of polymer chaperone-assisted translocation). Following Krapivsky and Mallick [353], consider a (rigid) polymer chain that passes through a membrane nanopore as shown in Fig. 7.41. Take the pore to be located at $x = 0$ and focus on the polymer segment to the right of the pore. At any given time t, the segment consists of L monomer units, each of size a. In the absence of any chaperones, the polymer executes an unbiased random walk, hopping by one monomer unit to the right or left at equal rates α. In the following we set $\alpha = a = 1$. Now suppose that the region on the right side of the membrane has a fixed density of chaperones that absorb irreversibly at a rate λ onto unoccupied monomeric sites of the polymer. A chaperone is assumed to be larger than the pore, so that it rectifies

the polymer diffusion when bound to the site immediately to the right of the pore (see also Sect. 4.1.2), resulting in a nonzero speed V.

(a) Let E_m denote the probability that an interval of length m to the right of the pore is chaperone-free. It follows that the mean speed of translocation is $V = 1 - E_1$, where $1 - E_1$ is the probability that there is a chaperone immediately to the right of the pore. Define the segment probability $S_m = E_m - E_{m+1}$, which is the probability that the leftmost chaperone is at a distance $m+1$ from the pore. The probabilities E_m then evolve according to

$$\frac{dE_m}{dt} = S_{m-1} - S_m - \lambda m E_m, \quad m \geq 1.$$

Using the definitions of E_m and S_m and the scenarios shown in Fig. 7.41, explain the meaning of each term on the right-hand side of this equation.

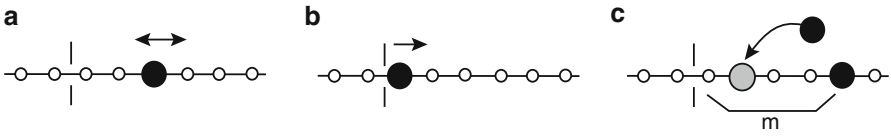


Fig. 7.41: Schematic of chaperone-assisted translocation model of [353]. (**a**) If the site adjacent to pore is unoccupied, then the polymer can hop in either direction. (**b**) The polymer can hop only to the right because an adsorbed chaperone is next to the pore and is too large to enter. (**c**) Adsorption of a new chaperone (*shaded*) at a site on the leftmost chaperone-free segment of length m

(b) Using the definition of S_m and the normalization condition $\sum_{m\geq 0} S_m = 1$ show that

$$\frac{dE_m}{dt} = E_{m-1} + E_{m+1} - 2E_m - \lambda m E_m, \quad m \geq 1,$$

with boundary condition $E_0 = 1$.

(c) Using the well-known identity of Bessel functions,

$$J_{\nu-1}(x) + J_{\nu+1}(x) - \frac{2\nu}{x} J_\nu(x) = 0,$$

and the boundary condition, show that the steady-state probabilities are given by

$$E_m = \frac{J_{m+2/\lambda}(2/\lambda)}{J_{2/\lambda}(2/\lambda)}$$

and hence that the expected speed is

$$V = 1 - \frac{J_{1+2/\lambda}(2/\lambda)}{J_{2/\lambda}(2/\lambda)}.$$

Plot V as a function of the absorption rate λ.

Problem 7.7 (Three-state model of bidirectional motor transport). Consider a three-state model of bidirectional motor transport

$$\partial_t p_1 = -v\partial_x p_1 + \frac{1}{\varepsilon}[-\beta_+ p_1 + \alpha p_3]$$
$$\partial_t p_2 = v\partial_x p_2 + \frac{1}{\varepsilon}[-\beta_- p_2 + \alpha p_3]$$
$$\partial_t p_3 = \frac{1}{\varepsilon}[\beta_+ p_1 + \beta_- p_2 - 2\alpha p_3].$$

Use the QSS reduction method to derive the FP equation

$$\frac{\partial C}{\partial t} = -V\frac{\partial C}{\partial x} + D\frac{\partial^2 C}{\partial x^2}$$

with

$$V = \frac{v}{\gamma}\left(\frac{1}{\beta_+} - \frac{1}{\beta_-}\right), \quad \gamma = \frac{1}{\beta_+} + \frac{1}{\beta_-} + \frac{1}{\alpha}$$

and

$$D = \varepsilon\left(\frac{(v-V)^2}{\gamma\beta_+^2} + \frac{(v+V)^2}{\gamma\beta_-^2}\right).$$

Hint: $w_0(x,n,t)$, $n = 1,2,3$, can be calculated explicitly using Gaussian elimination and the condition $\sum_n w_0(x,n,t) = 0$.

Problem 7.8 (Traveling wave approximation of a shock). Construct a traveling wave solution $\rho(x,t) = U(z)$ with $z = (x-ct)/\varepsilon$ of the equation

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho(1-\rho)) = \varepsilon\frac{\partial^2 \rho}{\partial x^2},$$

for $-\infty < x < \infty$ and wavespeed c. Assume that $U(z) \to U_{\pm\infty}$ as $z \to \pm\infty$. Show that

$$\frac{dU}{dz} = U(1-U) - cU + \text{constant},$$

and deduce that

$$c = \frac{U(1-U)]_{-\infty}^{\infty}}{[U]_{-\infty}^{\infty}}.$$

Use phase-pane analysis so that U can only tend to $U(\pm\infty)$ as $z \to \pm\infty$ if $dU/dz < 0$. Sketch a traveling wave solution and discuss how it relates to a shock solution of the kinematic wave equation obtained by setting $\varepsilon = 0$.

Problem 7.9 (Method of characteristics and shock waves). The kinematic wave equation arises in a wide range of transport applications including vehicular traffic. Suppose that $\rho(x,t)$ is the number density of cars on a single-lane road, evolving according to the equation

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho V(\rho))}{\partial x} = 0,$$

with $V(\rho)$ the density-dependent car speed. Take

$$V(\rho) = v_m(1-\rho/\rho_m),$$

where v_m is the speed limit and ρ_m is the maximum (bumper-to-bumper) car density. Suppose that the initial density profile is

$$\rho(x,0) = \begin{cases} \rho_m/8, & x<0 \\ \rho_m & x>0. \end{cases}$$

This represents cars on the left moving with speed $V = 7v_m/8$ encountering a traffic jam at $x = 0$. Use the method of characteristics to determine the density profile as a function of time. In particular, show that there is a shock that propagates from $x = 0$ at speed $\dot{s} = -v_m/8$. This represents a back-propagating shock that represents the slowing down of cars in response to the traffic jam ahead. Sketch the characteristics and shock in the $t-x$ plane.

Problem 7.10 (Model of DNA–protein interactions). Consider the model of DNA–protein interactions with the space-averaged FPT density $F(t)$ given by

$$F(t) = \sum_{n=1}^{\infty} \int_0^{\infty} dt_1 \ldots dt_n d\tau_1 \ldots d\tau_{n-1} \\ \delta\left(\sum_{i=1}^{n} t_i + \sum_{i=1}^{n-1} \tau_i - t\right) Q_1(t_n) \prod_{i=1}^{n-1} P_3(\tau_i) \prod_{i=1}^{n-1} P_1(t_i).$$

Let $\mathscr{L}(g) = \tilde{g}$ denote the Laplace transform of a function $g(t)$.

(a) Show that

$$\mathscr{L}(Q_1)(s) = \tilde{f}(s+\lambda_1), \quad \mathscr{L}(P_3)(s) = \frac{\lambda_3}{s+\lambda_3}, \quad \mathscr{L}(P_1)(s) = \lambda_1 \frac{1-\tilde{f}(s+\lambda_1)}{s+\lambda_1}.$$

(b) Laplace transforming $F(t)$ and summing the resulting geometric series obtain the result

$$\tilde{F}(s) = \tilde{f}(\lambda_1+s)\left[1 - \frac{1-\tilde{f}(\lambda_1+s)}{(1+s/\lambda_1)(1+s/\lambda_3)}\right]^{-1}.$$

(c) Determine the Laplace transform of the x-dependent FPT density $f(x,t)$ by solving

$$s\tilde{f}(x,s) = D_1 \frac{\partial^2 \tilde{f}(x,s)}{\partial x^2}, \quad 0 \le x \le L,$$

with boundary conditions $\tilde{f}(0,s) = 1$ and $\partial_x \tilde{f}(L,s) = 0$. Repeat for the case $-M \le x \le$ with boundary conditions $\tilde{f}(0,s) = 1$ and $\partial_x \tilde{f}(-M,s) = 0$.

(d) Averaging $\tilde{f}(x,s)$ with respect to x using the results of (c), show that

$$\tilde{f}(s) = \frac{1}{L+M}\sqrt{\frac{D_1}{s}}\left[\tanh(L\sqrt{s/D_1}) + \tanh(M\sqrt{s/D_1})\right].$$

(e) Finally, setting $\tau_i = 1/\lambda_i$, $i = 1,3$, derive the result

$$\begin{aligned}\tau &\equiv \int_0^\infty tF(t)dt = -\left.\frac{d\tilde{F}(s)}{ds}\right|_{s=0} \\ &= \left[\frac{(L+M)/\sqrt{\tau_1 D_1}}{\tanh(L/\sqrt{\tau_1 D_1}) + \tanh(M/\sqrt{\tau_1 D_1})} - 1\right](\tau_1+\tau_3).\end{aligned}$$

Problem 7.11 (Calculation of MFPT for unbiased random intermittent search). Consider the unbiased random search process given by Eq. (3.6.1). The MFPTs $T_\pm(y)$ to find the target, given that the particle starts at position y and state $\pm$ at time t, satisfy the pair of equations [see Eq. (7.6.21a, b)]

$$\begin{aligned}\partial_y T_+(y) + \frac{\beta}{v}[(\alpha u(y)-1)T_+(y) + \alpha u(y)T_-(y)] &= -\frac{\beta}{v}\left(\frac{1}{\beta}+u(y)\right), \\ \partial_y T_-(y) - \frac{\beta}{v}[\alpha u(y)T_+(y) + (\alpha u(y)-1)T_-(y)] &= \frac{\beta}{v}\left(\frac{1}{\beta}+u(y)\right).\end{aligned}$$

(a) Transform the equations using the new variables

$$S_1(y) = \frac{T_+(y)+T_-(y)}{2}, \quad S_2(y) = \frac{T_+(y)-T_-(y)}{2}.$$

(b) Solve the equations for $S_{1,2}(y)$ in each of the following regions: (I) $-L < y < -a$, (II) $-a < y < a$, and (III) $a < y < L$. It is necessary to use the boundary conditions $D(\pm L) = 0$ and to impose continuity at $y = \pm a$.

(c) Show that $S(y) = S(-y)$ and hence that the average MFPT is given by

$$\begin{aligned}\tau_1 &= \frac{1}{L}\int_0^L S(y)dt \\ &= \frac{1}{L}\left(\frac{1}{\beta}+\frac{1}{2\alpha}\right)\left(\left[\frac{\beta}{v}\right]^2\frac{2(L-a)^3}{3} + \frac{\beta}{v}\sqrt{\frac{2\alpha+k}{k}}(L-a)^2\coth(\Lambda a)\right) \\ &\quad + \frac{2\alpha+k}{k}\left(\frac{1}{\beta}+\frac{1}{2\alpha+k}\right)\frac{L-a}{L},\end{aligned}$$

where

$$\Lambda = \frac{\beta}{v}\sqrt{\frac{k}{2\alpha+k}}.$$

Problem 7.12 (Computer simulations: random intermittent searcher in 1D).

(a) Write a computer program to simulate an unbiased three-state random intermittent searcher moving along a 1D track of length $L = 20\,\mu\text{m}$. Take the velocity to be $v = 0.1\,\mu\text{m}\,\text{s}^{-1}$, and the transition rates $\beta = 1\,\text{s}^{-1}, \alpha = 0.5\,\text{s}^{-1}$. Assume that there is a target of width $a = 1\,\mu\text{m}$ at position $X = 10\,\mu\text{m}$, and the particle finds (is absorbed by) the target at a rate $k = 0.05\,\text{s}^{-1}$. (Hint: use the Gillespie algorithm to determine the random transition times and the sequence of states while keeping track of changes in position. Hence, if the particle enters the $+$ $(-)$ state at $t = t_0$ and makes the next transition at time $t_0 + \tau$, then the position is shifted by an amount $\Delta x = v\tau$ ($\Delta x = -v\tau$). There is no shift if the particle is in the stationary state.)

(b) Plot sample trajectories up to the time T the target is found. By averaging T over many trials determine the MFPT to find the target, starting from a random initial position. Compare with the analytical expression obtained in part (c) of Ex. 7.11.

(c) Modify the program so that it now simulates a particle executing an unbiased 1D random walk along the track for a random time τ_{1D}, after which it is removed from the track and placed randomly at a new location on the track after a random time τ_{3D}. Assume that the times τ_j, $j = 1D, 3D$, are exponentially distributed with means $\bar{\tau}_j$, and select the new position using the uniform distribution on $[0, L]$. Take $\bar{\tau}_{3D} = \bar{\tau}_{1D} = 10^{-3}\,\text{s}$ and 1D diffusivity $D_1 = 10\,\mu\text{m}^2\,\text{s}^{-1}$.

Chapter 8
Self-Organization in Cells I: Active Processes

Another fundamental question in modern cell biology is how cellular and subcellular structures are formed and maintained given their particular molecular components. How are the different shapes, sizes, and functions of cellular organelles determined, and why are specific structures formed at particular locations and stages of the life cycle of a cell? In order to address these questions it is necessary to consider the theory of self-organizing non-equilibrium systems [441]. One system where the principles of self-organization are widely applied is the cytoskeleton [275, 442, 467]. The cytoskeleton, which is made up of polymers such as microtubules and F-actin, determines cell shape and polarity, maintains the structural integrity of the cell, and forms the mitotic spindle apparatus during cell division. As described in Sect. 4.1, F-actin and microtubules are polarized filaments that are intrinsically unstable, undergoing continuous turnover of subunits by addition at their plus end and depolymerization at their minus end. The continuous dynamic exchange of the subunits and their interactions with filament-associated proteins provides the basis for the self-organization of different cytoskeletal structures. This has been demonstrated in vitro for microtubule networks, which form by simply combining tubulin, microtubule motors, and ATP in solution [467]. By varying the relative concentrations of motors and tubulin, one can generate different network patterns, including random networks, vortices, or asters.

In this chapter we focus on the role of polymerization in the self-organization of cytoskeletal structures. Reaction–diffusion models of cellular self-organization will be considered in Chap. 9. We begin by considering various mechanisms for cellular length regulation, based on the interactions between polymerizing filaments and molecular motors (Sect. 8.1). Self-organization of the mitotic spindle during various stages of cell mitosis is considered in Sect. 8.2. We describe the search-and-capture model for the interactions between chromosomes and the mitotic spindle, dynamical instabilities in the positioning of the chromosomes, and spindle length control. In Sect. 8.3 we discuss the important role of actin polymerization and Brownian ratchets in cell motility. Finally, we briefly consider a different example of an active self-organizing process based on vesicular transport between distinct intracellular compartments (Sect. 8.4).

P.C. Bressloff, *Stochastic Processes in Cell Biology*, Interdisciplinary Applied Mathematics 41, DOI 10.1007/978-3-319-08488-6_8

8.1 Cellular Length Regulation by Molecular Motors

One basic issue is how the sizes of subcellular structures are determined in order to scale with the size of the cell and with physiological requirements. It appears that self-organizing processes together with physical constraints play a major role in controlling organelle size [519]. At least three distinct control mechanisms have been identified.

1. *Molecular rulers.* In the case of linear structures such as filaments, size control can be achieved by a molecular ruler protein, whose length is equal to the desired length of the growing structure. One classical example is the length of the λ-phage tail, which is determined by the size of the gene H product (gpH) [319]. During assembly of the tail, gpH is attached to the growing end in a folded state, and protects the growing end from the terminator gene product U (gpU). As the tail elongates, gpH stretches such that when it is fully extended, further growth exposes the tail to the action of gpU, (see Fig. 8.1).

2. *Quantal synthesis.* Size could be controlled by synthesizing exactly enough material to build a structure of the appropriate size—a process known as quantal synthesis. For example, precursor protein levels are known to affect the length of flagella in the unicellular green alga *Chlamydomonas reinhardtii* [378], and the length of sea urchin cilia is correlated with the concentration of the protein tektin [608]. One prediction of the quantal synthesis model is that doubling the number of flagella should halve their length. However, studies of *Chlamydomonas* mutants indicate a much weaker dependence of length on the number of flagella, suggesting that there is an additional length-controlling mechanism involving dynamic balance [420]; see below. Another example of organelle size regulation via protein synthesis occurs for the endoplasmic reticulum (ER). In this case, there appears to be a feedback signaling pathway, the unfolded protein response (UPR), which monitors the state of the ER and controls the synthesis of membrane lipids [129].

3. *Dynamic balance.* Dynamic structures are constantly turning over so that in order for them to maintain a fixed size, there must be a balance between the rates of assembly and disassembly. If these rates depend on the size in an appropriate way,

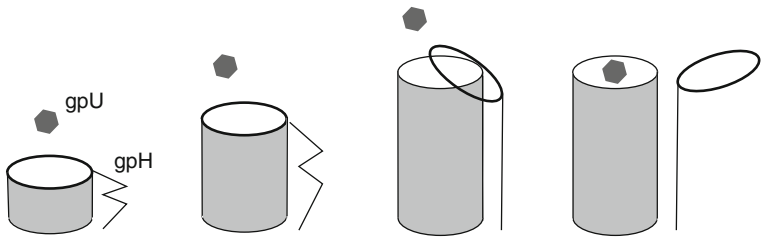


Fig. 8.1: Mechanism of cellular size control: Molecular ruler in bacteriophage tail

then there will be a unique balance point that stabilizes the size of the organelle. For example, eukaryotic flagellar microtubules undergo continuous assembly and disassembly at their tips, in which a constant rate of disassembly is balanced by a length-dependent rate of assembly due to a fixed number of molecular motors transporting tubulin dimers from the cell body, leading to a fixed flagellar length [419, 420]. The transport of tubulin is also thought to be one possible mechanism for controlling axonal elongation. In this case diffusion, possibly combined with an active component, transports newly synthesized tubulin at the somatic end of the axon to the + end of microtubules at the axonal tip. As the axon grows, the concentration of tubulin at the tip decreases until there is a balance between polymerization and depolymerization. (A simple continuum model of axonal elongation [431] is considered in Ex. 8.1. For a more detailed stochastic model see [15].) A different balance mechanism appears to control the length of microtubules in yeast, where kinesin motors move processively to the microtubule tips where they catalyze disassembly. Longer microtubules recruit more kinesin motors from the cytoplasm, which results in a length-dependent rate of disassembly. When this is combined with a length-independent rate of assembly, a unique steady-state microtubule length is obtained [652]. In more complex cytoskeletal structures such as the mitotic spindle (Sect. 8.2), it is likely that multiple length-dependent mechanisms play a role [84]. Dynamic balance mechanisms also control other filament-like structures. For example, actin filaments within stereocilia of the inner ear [549], (see Sect. 5.4), constantly treadmill back towards the cell body, with disassembly at the base balanced by assembly at the tip. The latter depends on the diffusion of actin monomers to the tip, which results in a length-dependent rate of assembly. Yet another example is the control of the hook length in bacterial flagella [320].

In this section we consider cellular length-control mechanisms based on the action of molecular motors. There have been a number of theoretical models that combine the dynamics of molecular motors (Sect. 4.4) with the dynamics of microtubule assembly and disassembly (Sect. 4.1). Several studies have focused on motor regulation of depolymerization at the tip combined with constant growth [234, 274, 435, 525] or treadmilling [304], whereas others have considered the effects of motors on dynamic instabilities, that is, the frequency of catastrophes [357, 629]. We will describe simplified versions of these models developed in [274, 357]. The more complicated example of mitotic spindle length control during cell division will be considered in Sect. 8.2.3.

8.1.1 Filament Length Regulation by Depolymerization

Let $\rho(x,t)$ denote the density of molecular motors along a single filament, which is assumed to evolve according to the mean-field model (see also equation (7.5.4)) [357, 496]

$$\frac{\partial \rho}{\partial t} = -v\frac{\partial}{\partial x}\left[\rho(1-\rho/\rho_{\max})\right] + k_{\rm on}c(1-\rho/\rho_{\max}) - k_{\rm off}\rho. \tag{8.1.1}$$

Here v is the speed of the motors in the absence of crowding, $\rho_{\max}$ is the maximum possible motor density, and the factors $(1-\rho/\rho_{\max})$ take into account the effects of molecular crowding on the processivity of the motors and on their rate of binding to the filament. The latter is proportional to the bulk motor concentration c in the cytoplasm, which is assumed fixed. As a further simplification, the effects of crowding on the drift term are neglected so that the PDE is linear. Introducing the fractional motor density $p(x,t) = \rho(x,t)/\rho_{\max}$, we have

$$\frac{\partial p}{\partial t} = -v\frac{\partial p}{\partial x} + \bar{k}_{\rm on}(1-p) - k_{\rm off}p, \quad \bar{k}_{\rm on} = \frac{k_{\rm on}c}{\rho_{\max}}. \tag{8.1.2}$$

Equation (8.1.2) is supplemented by the boundary condition $p(0,t) = 0$ (absorbing boundary condition at the base of the filament) and the initial condition $p(x,0) = 0$. Away from the tip of the microtubule, it is assumed that the motor density reaches a QSS p_s given by the solution to the equation

$$\frac{dp_s}{\partial x} + \frac{1}{\lambda}p_s = \frac{\bar{k}_{\rm on}}{v}, \quad \lambda = \frac{v}{k_{\rm off}+\bar{k}_{\rm on}}$$

Hence,

$$p_s(x) = p_0\left[1-{\rm e}^{-x/\lambda}\right], \quad p_0 = \frac{\bar{k}_{\rm on}\lambda}{v} = \frac{\bar{k}_{\rm on}}{k_{\rm off}+\bar{k}_{\rm on}}.$$

Following [274], suppose that the depolymerization rate at the tip is $\gamma p_e(t)$, where $p_e(t)$ is the fractional motor density at the tip and γ is a constant. Taking the rate of growth to be a constant α, the filament length $L(t)$ evolves as

$$\frac{dL}{dt} = \alpha - \gamma p_e. \tag{8.1.3}$$

Let a be the lattice spacing of a single subunit of the filament. The rate at which molecular motors enter the tip (assuming the site is unoccupied) is then $(v-L')/a$ with $L' = dL/dt$. It follows that p_e satisfies the equation

$$\frac{dp_e}{dt} = \frac{v-L'}{a}p(L-a,t)(1-p_e) - k_{\rm e,off}p_e, \tag{8.1.4}$$

where $k_{\rm e,off}$ is the rate of unbinding at the tip. The quasi-static assumption implies that

$$p(L-a,t) = p_s(L-a) \approx p_0(1-{\rm e}^{-L/\lambda}).$$

Therefore, at steady state ($p_e' = L' = 0$), equation (8.1.4) implies that $p_e = p_e(L)$ with

$$p_e(L) = \frac{p_0(1-{\rm e}^{-L/\lambda})}{ak_{\rm e,off}/v + p_0(1-{\rm e}^{-L/\lambda})},$$

which can be rearranged to give

$$1 - \mathrm{e}^{-L/\lambda} = \frac{a k_{\mathrm{e,off}}}{p_0 v} \frac{p_e}{1 - p_e}.$$

However, we also have $p_e = \alpha/\gamma$, which then determines the filament length to be

$$L_s = -\lambda \ln \left[1 - \frac{a k_{\mathrm{e,off}}}{p_0 v} \frac{\alpha}{\gamma - \alpha} \right]. \tag{8.1.5}$$

The existence of a steady-state length requires the argument of the logarithm to be positive, which means that

$$p_0 > p_{0c} = \frac{\alpha k_{\mathrm{e,off}} a}{v(\gamma - \alpha)}.$$

Note that p_0 is a monotonically increasing function of the bulk motor concentration c. Hence, for a steady-state length to occur the bulk motor concentration must exceed a minimal concentration c_0 with

$$\frac{k_{\mathrm{on}} c_0}{k_{\mathrm{off}} \rho_{\max} + k_{\mathrm{on}} c_0} = \frac{a k_{\mathrm{e,off}}}{v(\gamma/\alpha - 1)},$$

which yields

$$c_0 = \frac{k_{\mathrm{off}} k_{\mathrm{e,off}} \rho_{\max} a}{k_{\mathrm{on}}} \frac{1}{v(\gamma/\alpha - 1) - a k_{\mathrm{e,off}}}. \tag{8.1.6}$$

Note that L_s is a monotonically decreasing function of c with $L_s \to \infty$ as $c \to c_0$ from above.

The above model ignores statistical fluctuations. However, Monte Carlo simulations of a stochastic version of the model show that there is a unimodal distribution of filament lengths that is peaked around the steady-state value of mean-field theory

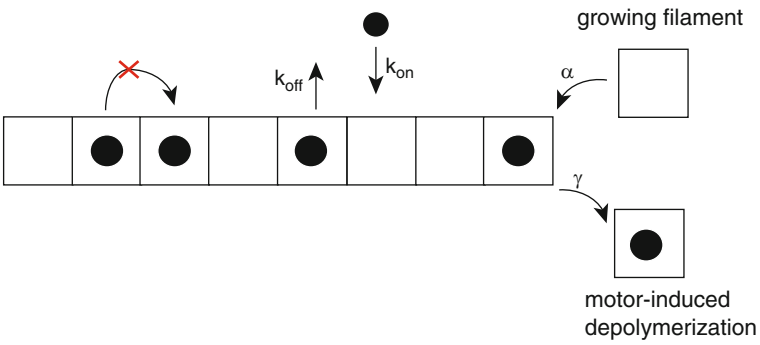


Fig. 8.2: Schematic diagram of filament length regulation by molecular motor-based depolymerization

[274, 357]. The stochastic model takes the form of a TASEP in which a bound motor can hop to a neighboring site if it is unoccupied, (see Fig. 8.2). Similarly a motor in the bulk can only bind to a site if it is unoccupied. A monomer can be removed from the tip of the filament when it is occupied by a motor. Following removal of the monomer, the motor steps backward if the penultimate site is unoccupied; otherwise it is also removed. The tip motor can also spontaneously unbind without removal of the end monomer. The results from computer simulations are shown in Fig. 8.3 for both the density-controlled model and a flux-controlled model. The latter takes the depolymerization rate to depend on the flux of motors at the tip, rather than the motor density. That is [357],

$$L' = \alpha - a\rho_{\max} p_0 \left[1 - \mathrm{e}^{-L/\lambda}\right](v - L'),$$

and the steady-state length is

$$L_s = -\lambda \ln\left[1 - \frac{\alpha}{p_0 v}\delta\rho_{\max}\right]. \tag{8.1.7}$$

One finds statistical fluctuations are reduced in the flux-controlled model.

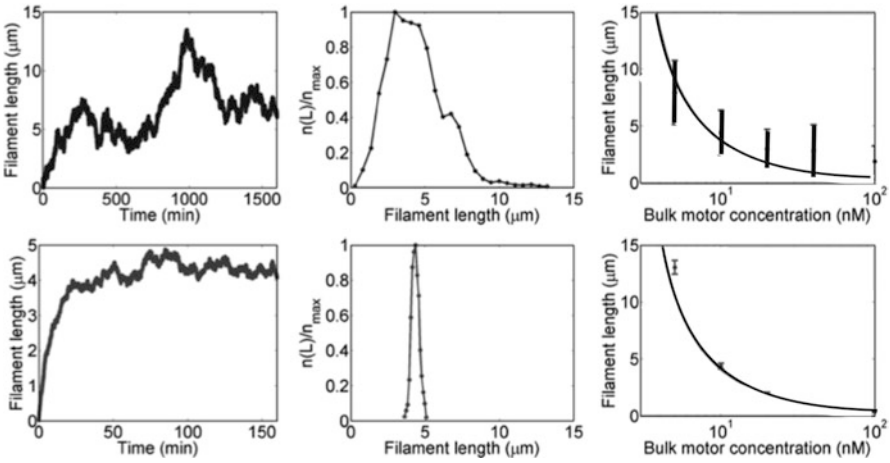


Fig. 8.3: Filament dynamics and steady-state filament length for length regulation by depolymerization. *Top row*: density-controlled depolymerization. *Bottom row*: flux-controlled depolymerization. Left: example trace of filament length versus time from a simulation of TASEP model. *Middle*: normalized filament length distribution averaged with respect to ten stochastic simulations (after removal of initial transients). *Right*: comparison of steady-state filament length based on mean-field theory with stochastic simulation (error bars are standard deviations of steady-state length distributions). Parameters of mean-field model are $v = 3\ \mu\mathrm{m\,min}^{-1}$, $k_{\mathrm{on}} = 2\ \mathrm{nM}^{-1}\mu\mathrm{m}^{-1}\mathrm{min}^{-1}$, $k_{\mathrm{off}} = 0.25\ \mathrm{min}^{-1}$, $k_{\mathrm{e,off}} = 1.45\ \mathrm{min}^{-1}$, $\gamma = 1.025\ \mu\mathrm{m\,min}^{-1}$, $a = 8$ nm, $\delta = 8$ nm, and $\rho_{\max} = 125\ \mu\mathrm{m}^{-1}$. For the density-controlled model $\alpha = 1.0\ \mu\mathrm{m\,min}^{-1}$, while for the flux-controlled model $\alpha = 0.5\ \mu\mathrm{m\,min}^{-1}$. The stochastic simulations use the same parameters except $\gamma = 1.5\ \mu\mathrm{m\,min}^{-1}$ and $k_{\mathrm{e,off}} = 1\ \mathrm{min}^{-1}$ for the density-controlled model (Adapted from Kuan and Betterton [357])

8.1.2 Filament Length Regulation by Altering the Frequency of Catastrophes

Some experimental and modeling studies have suggested that, rather than directly regulating depolymerization, kinesin motors promote the transition from growing to shrinking phases (catastrophes) in dynamic microtubules [208, 209, 248, 357, 629] (see Fig. 8.4). Kuan et al. [357] consider a modified version of the Dogterom–Leibler model of microtubule catastrophes [146] (see also equation (4.1.10))

$$\frac{\partial n_+}{\partial t} = -v_+ \frac{\partial n_+}{\partial L} - k_+ n_+ + k_- n_- \tag{8.1.8a}$$

$$\frac{\partial n_-}{\partial t} = v_- \frac{\partial n_-}{\partial L} - k_- n_- + k_+ n_+, \tag{8.1.8b}$$

in which the transition rate k_+ from the growing to shrinking phase is taken to depend on the fractional density p_e of motors at the tip:

$$k_+ = \bar{k}_+ + \alpha p_e.$$

Here $n_\pm(L,t)$ represent the number density of filaments of length L in the growing and shrinking phase, respectively. In steady state (assuming it exists), the transition rate k_+ will be L-dependent due to the L-dependence of p_e. From the steady-state

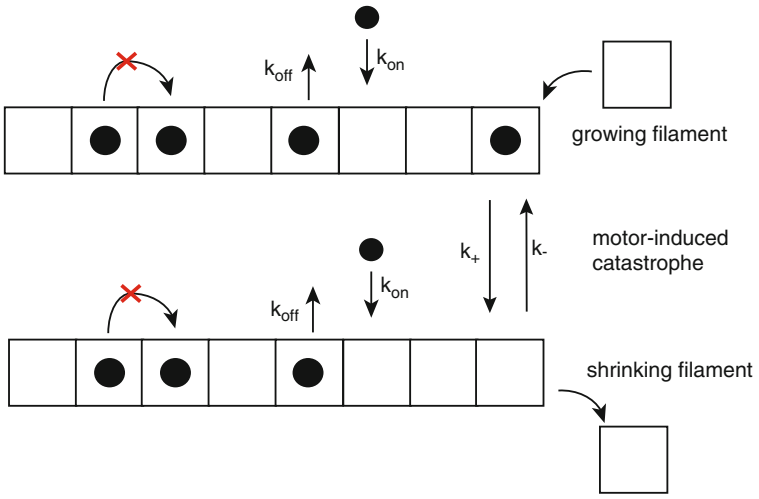


Fig. 8.4: Schematic diagram of filament length regulation by altering the frequency of catastrophes

analysis of equation (4.1.10) in Sect. 4.1, it is straightforward to show that the number density of filaments of length L is given by $n_\pm(L) = n(L)/v_\pm$ with

$$n(L) = n(0) \mathrm{e}^{k_- L/v_-} \exp\left(-\frac{1}{v_+} \int^L k_+(L')dL' \right). \tag{8.1.9}$$

The length dependence of k_+ is obtained by solving equation (8.1.4) with $p_e' = 0$ and $L' = v_+$, the speed of the growing phase. Hence,

$$p_e = p_e(L) = \frac{p_0(1 - \mathrm{e}^{-L/\lambda})}{ak_{\mathrm{e,off}}/(v - v_+) + p_0(1 - \mathrm{e}^{-L/\lambda})},$$

and $k_+(L) = \bar{k}_+ + \alpha p_e(L)$. If $\mathrm{e}^{-L/\lambda} \ll 1$, then we can neglect the L-dependence of k_+ and [357]

$$n(L) \approx n(0)\mathrm{e}^{-[(\bar{k}_+ + \Delta k_+)/v_+ - k_-/v_-]L}, \tag{8.1.10}$$

with

$$\Delta k_+ = \frac{\alpha(v - v_+)p_0}{ak_{\mathrm{e,off}} + (v - v_+)p_0}.$$

It follows that increasing p_0 by increasing the bulk motor concentration results in a higher catastrophe rate and, hence, shorter filament lengths.

8.1.3 Length Regulation by Intraflagellar Transport

Radioactive pulse labeling has been used to measure protein turnover in eukaryotic flagella. Such measurements have established that turnover of tubulin occurs at the + end of flagellar microtubules and that the assembly (rather than disassembly) part of the turnover is mediated by intraflagellar transport (IFT). This is a motor-

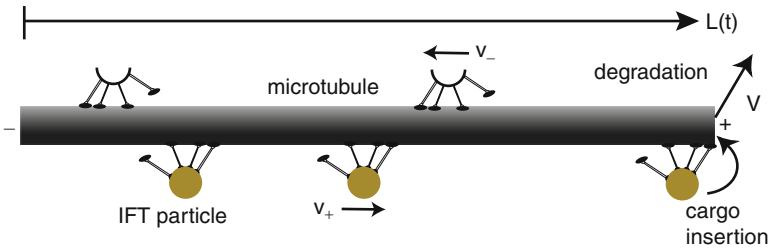


Fig. 8.5: Schematic diagram of intraflagellar transport (IFT), in which IFT particles travel with speed $v_\pm$ to the $\pm$ end of a flagellum. When an IFT particle reaches the + end it releases its cargo of protein precursors that contribute to the assembly of the flagellum. Disassembly occurs independently of IFT transport at a speed V

assisted motility within flagella in which large protein complexes move from one end of the flagellum to the other [572]. Particles of various size travel to the flagellar tip (anterograde transport) at 2.0 μm/s, and smaller particles return from the tip (retrograde transport) at 3.5 μm/s after dropping off their cargo of assembly proteins at the + end. A schematic diagram of IFT transport is shown in Fig. 8.5. Immunofluorescence analysis indicates that the number of IFT particles (estimated

to be in the range 1–10) is independent of length [419, 420]. If a fixed number of transport complexes M move at a fixed mean speed $\bar{v}$, then the rate of transport and assembly should decrease inversely with the flagellar length L. On the other hand, measurements of the rate of flagellar shrinkage when IFT is blocked indicate that the rate of disassembly is length-independent. This has motivated the following simple deterministic model for length control [419]:

$$\frac{dL}{dt} = \frac{a\bar{v}M}{2L} - V, \tag{8.1.11}$$

where a is the size of the precursor protein transported by each IFT particle and V is the speed of disassembly. Equation (8.1.11) has a unique stable equilibrium given by $L^* = a\bar{v}M/2V$. Using the experimentally based values $M = 10$, $\bar{v} = 2.5$ μm/s, $L^* = 10$ μm, and $V = 0.01$ μm/s, the effective precursor protein size is estimated to be $a \approx 10$ nm. (A stochastic version of a model for IFT has also been developed using the theory of continuous-time random walks [64]; see Ex. 8.2.)

8.2 Cell Mitosis

Mitosis is a phase in a eukaryotic cell's life cycle, during which it segregates its already-duplicated chromosomes in preparation for cell division, or *cytokinesis*. Following duplication of its DNA, a cell synthesizes many additional macromolecules, so at the end of the period between cell divisions (*interphase*), all of the materials needed to form two viable cells are present. Mitosis and cytokinesis then separate this biochemically doubled cell into two essentially identical objects, each equipped to grow and divide again. The major molecular machinery responsible for organizing and segregating the duplicated chromosomes is known as the *mitotic spindle*, which is an assembly of microtubules (MTs) spreading radially from two poles. Mitosis consists of several distinct phases as illustrated in Fig. 8.6 [428]. An image of a mitotic spindle obtained using fluorescent microscopy (see also Sect. 1.2) is shown in Fig. 8.7.

(i) *Prophase.* The first physical step known as *prophase* involves restructuring the chromosomes within a dividing cell, so that each is sufficiently compact to be separable within a space no bigger than a single cell. In a human cell, the DNA molecules range in length from 1.9 to 8.5 cm, whereas the nucleus that contains 46 copies is an approximately spherical compartment with diameter usually less than 8 mm. Thus, each DNA copy must be reduced in length by more than a factor of 1,000. This form of condensation is achieved in multiple steps. First, the DNA is wrapped around *nucleosome* "core particles" resulting in the material *chromatin*. Fibers of chromatin are then somehow coiled and looped until each chromosome is only a few micrometers long and less than 1 mm thick. As condensation proceeds the pair of sister DNA duplexes become structurally distinct. These so-called chromatids remain tightly coupled around their central region known as the *centromere*. Such coupling is crucial for the proper functioning of mitosis, since accurate segregation of sister chromatids

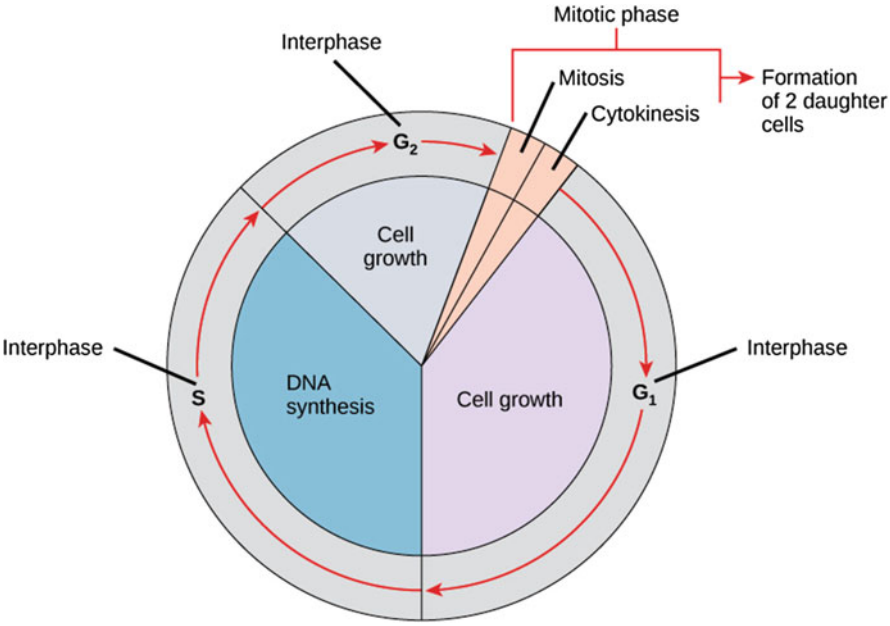


Fig. 8.6: The cell cycle of eukaryotes consists of four distinct phases or stages: G1, S, G2 and M. Here G1 and G2 are *gap* phases, while S is the *synthesis* stage where DNA replication occurs. The M phase consists of two tightly coupled processes; *mitosis* in which the cell's chromosomes are divided into two and *cytokinesis* where the cell's cytoplasm divides in half to form distinct cells. Activation of each phase is dependent on the proper progression and completion of the previous one. Cells that have temporarily stopped dividing are said to have entered a state of quiescence called the G0 phase. The mitotic phase is itself subdivided into four distinct phases, see text for details (Public domain figure downloaded from Boundless)

depends on their being attached until the moment when all chromatids simultaneously begin segregation. The centromere is also the domain where each chromosome develops specializations such as the *kinetochore*, which is the protein structure on chromatids where the mitotic spindle fibers attach during cell division to pull sister chromatids apart.

(ii) *Prometaphase.* Spindle formation initiates the process of chromosome organization during *prometaphase*. The main step is the attachment of all chromosomes to spindle MTs in such a way that each chromatid of every chromosome is associated with MTs that are in turn associated with one and only one end of the mitotic spindle. A second step of prometaphase is the migration of all chromosomes to the spindle mid-plane or equator, a process called *congression*.

(iii) *Metaphase.* Once the chromosomes are positioned at the equator, the cell is said to be in *metaphase*. Normal cells include quality control processes that determine whether each chromosome is properly attached to the spindle before

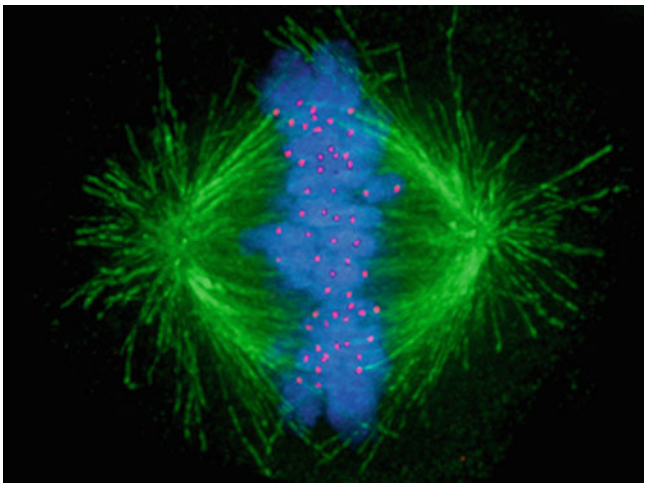


Fig. 8.7: Image of the mitotic spindle in a human cell showing microtubules in *green*, chromosomes (DNA) in *blue*, and kinetochores in *red* (Public domain figure downloaded from Wikipedia)

segregation is allowed to begin; this is the *spindle assembly checkpoint* (SAC). Shortly after this checkpoint has been satisfied, the cohesins that have been holding sister chromatids together are cleaved by a protease.

(iv) *Anaphase.* MT-generated forces acting on the now-independent sister chromatids move them to opposite ends of the cell in a process called *anaphase*. If the nuclear envelope dispersed during spindle formation, then it now reforms on the still-condensed chromosomes by the application of vesicles derived largely, if not entirely, from the previously dissociated nuclear envelope. As these membranes are fusing to define the two nuclear compartments, the cell initiates cytokinesis, the process that divides the cytoplasm into two approximately equal parts, each of which contains its own nucleus. At the same time, the chromosomes de-condense, and the daughter cells return to interphase.

A number of mathematical and computational models have been developed in order to gain an understanding of how the spindle machinery performs its complex functions. Typically, a model focuses on one or two aspects of mitosis such as spindle assembly, positioning, maintenance and elongation, chromosomal capture and congression, and the SAC. In the following we describe some of these models; see also the reviews by Mogilner et. al. [445, 450].

8.2.1 Search-and-Capture Model of Chromosome/Kinetochore Attachment

A crucial step in prometaphase is the attachment of each chromosome to a kinetochore of the mitotic spindle. According to the *search-and-capture* model of Kirschner and Mitchison [336], the underlying mechanism involves the nucleation of MTs in random directions, which then grow and shrink dynamically in order to search space and eventually encounter a target kinetochore (see Fig. 8.8). Analysis of a single MT searching for a single kinetochore shows that MT dynamic instability (Sect. 4.1.2) provides an effective search mechanism provided it is regulated appropriately, that is, MTs do not waste time growing in the wrong direction and don't undergo premature catastrophe when growing in the right direction [260, 272]. However, as highlighted by Wollman et al. [685], although the estimated capture time is consistent with the duration of the mitotic phase, it does not take into account realistic geometries nor the capture of multiple chromosomes. Using a combination of mathematical analysis and computer simulations, Wollman et al. show that unbiased search and capture for multiple chromosomes is not efficient enough to account for the duration of the prometaphase. On the other hand, if there exists a spatial gradient in some stabilizing factor that biases MT dynamics towards the chromosomes, then one obtains more realistic capture times [685]. One candidate molecule for acting as a stabilizing factor is Ran-GTP [95]; see also Sect. 9.1.

We now develop the analysis of unbiased search and capture, following along the lines of Wollman et al. [685], see also [229]. We begin by considering a single MT searching for a single kinetochore. Let $P(n)$ be the probability that n sequentially nucleated MTs fail to capture the kinetochore but the $(n+1)$th MT is successful. Let $P(\tau|n)$ denote the conditional probability that given n cycles of failure, the time to capture is less than τ. The total probability of capture before time τ is then

$$P(\tau) = \sum_{n=0}^{\infty} P(\tau|n)P(n). \tag{8.2.1}$$

The probability $P(n)$ is given by

$$P(n) = p(1-p)^n,$$

where p is the probability of an MT nucleating in the right direction and reaching the kinetochore before undergoing catastrophe. Suppose that the kinetochore is at a radial distance x from the nearest pole of the mitotic spindle and has an effective target radius of r. Assuming that the MTs are nucleated in random directions, it follows that the probability P_1 of an MT nucleating in the right direction is given by the solid angle subtended by the target:

$$P_1 = \frac{\pi r^2}{4\pi x^2} = \frac{r^2}{4x^2}.$$

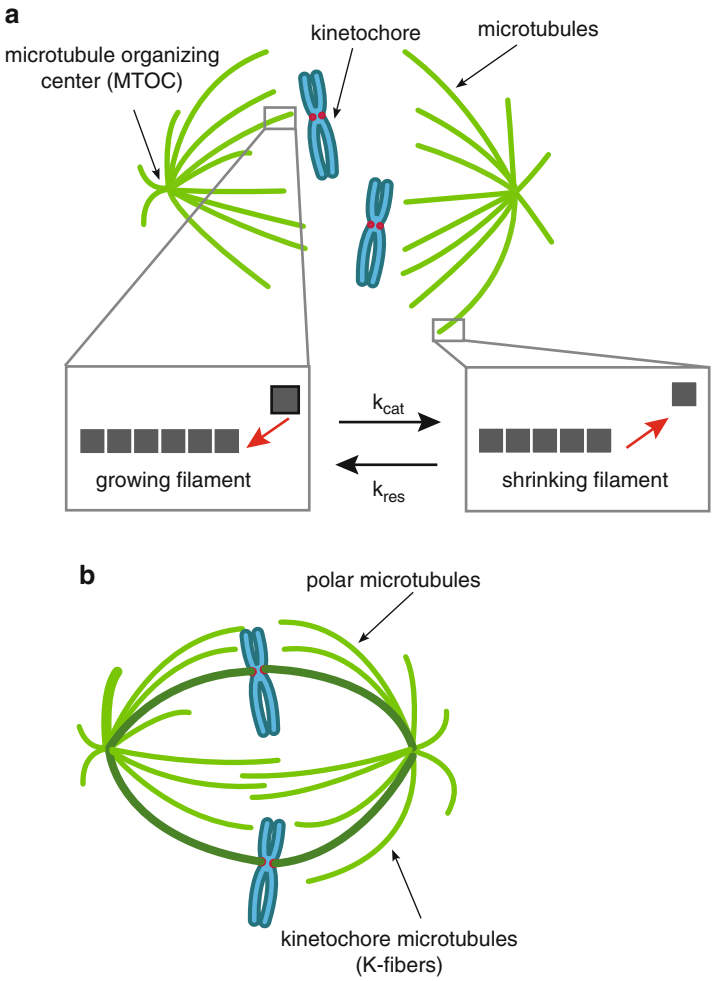


Fig. 8.8: Schematic diagram of search-and-capture model based on microtubule dynamic instability. (**a**) During prometaphase MTs randomly probe the cellular domain by alternating between growth and shrinkage phases until they capture the kinetochores. This process of dynamic instability can be quantified by four parameters: the rates of growth (V_g) and shortening (V_S) and the rates of catastrophe (k_{cat}) and rescue (k_{res}). To capture all kinetochores in a reasonable time frame, the dynamic instability parameters have to be optimized, but this is not sufficient to make the process fast enough. One possible mechanism for accelerating the search and capture is the presence of a RanGTP gradient around the chromosomes that biases the MT dynamics [685]. (**b**) At the end of prometaphase, all the kinetochores are attached to MTs, one from each pole of the mitotic spindle, and are co-aligned along the mid-plane

Suppose that the time from nucleation of an MT to its catastrophe is exponentially distributed. (For simplicity, it is assumed that following catastrophe an MT cannot be rescued, that is, it starts growing again.) This means that the probability P_2 of reaching the kinetochore before catastrophe is

$$P_2 = \int_{T_s}^{\infty} k_{\rm cat} \mathrm{e}^{-k_{\rm cat} t} dt = \mathrm{e}^{-k_{\rm cat} x/V_g},$$

where $T_s = x/V_g$ is the time to reach the kinetochore given a constant speed of growth V_g. Therefore, the probability of reaching the kinetochore is

$$p = P_1 P_2 = \frac{r^2}{4x^2} \mathrm{e}^{-x k_{\rm cat}/V_g}. \tag{8.2.2}$$

It remains to calculate the conditional probability $P(\tau|n)$. First, note that $P(\tau|n) = Q(\tau - \Delta\tau|n)$ where $\Delta\tau = x/V_g$ is the time for the $(n+1)$th MT to reach the kinetochore and $Q(\tau|n)$ is the conditional probability that given n cycles of failure ($n \geq 1$), the total time taken up by these cycles is less than τ. In order to calculate the latter, we need to determine the average lifetime $T_{\rm cycle}$ of an unsuccessful cycle, starting from nucleation through catastrophe to complete depolymerization. The mean time to a catastrophe is $T_c = 1/k_{\rm cat}$ and the subsequent time for the MT to shrink is given by the mean length at the start of catastrophe divided by the speed of shrinkage, $V_g/(V_s k_{\rm cat})$. Therefore,

$$T_{\rm cycle} = \frac{V_g + V_s}{V_s k_{\rm cat}}. \tag{8.2.3}$$

The duration of each nucleation cycle, in the absence of rescue, is an exponential random variable with mean $T_{\rm cycle}$. We now use the basic result that the sum of n exponential random variables is a Gamma random variable (see Box 8A), that is,

$$Q(\tau|n) = \frac{1}{T_{\rm cycle}^n (n-1)!} \int_0^{\tau} s^{n-1} \mathrm{e}^{-s/T_{\rm cycle}} ds. \tag{8.2.4}$$

We can now evaluate $P(\tau)$ according to (see Ex. 8.3)

$$\begin{aligned} P(\tau) &= p + \sum_{n=1}^{\infty} (1-p)^n p Q(\tau - \Delta\tau|n), \tau > \Delta\tau, \\ &= p + (1-p)(1 - \mathrm{e}^{-p(\tau - \Delta\tau)/T_{\rm cycle}}). \end{aligned}$$

If $p \ll 1$, then the characteristic number of unsuccessful searches is $n \gg 1$ and the typical search time is $\tau \gg \Delta t$. Hence

$$P(\tau) \approx 1 - \mathrm{e}^{-p\tau/T_{\rm cycle}},$$

which implies that the average time to capture is

$$T_{\text{capture}} = \frac{T_{\text{cycle}}}{p} = \frac{V_g + V_s}{V_s k_{\text{cat}}} \frac{4x^2}{r^2} e^{x k_{\text{cat}}/V_g}. \tag{8.2.5}$$

It follows that the optimal catastrophe frequency is $k_{\text{cat}} = V_g/x$ [272, 685]. Extensions to the case of multiple MTs and multiple kinetochores are considered in Ex 8.2. For example, one finds that for N nucleating MTs and a single kinetochore, the mean time to capture is T_{capture}/N. Moreover, for N MTs and M kinetochores ($N \geq M$), the mean time to capture all of the kinetochores is approximately

$$T_{N,M} = (T_{\text{capture}}/N) \ln M.$$

It turns out that the time to capture $M = 46$ chromosomes using up to 1,000 searching MTs is substantially greater than experimental measurements of 20–30 min. This was shown by Wollman et al. [685] using Monte Carlo computer simulations of their model based on the following algorithm:

1. Chromosome positions are generated randomly within a sphere of radius $10\mu m$ representing the nucleus, and for each chromosome the two pole-kinetochore distances are calculated.

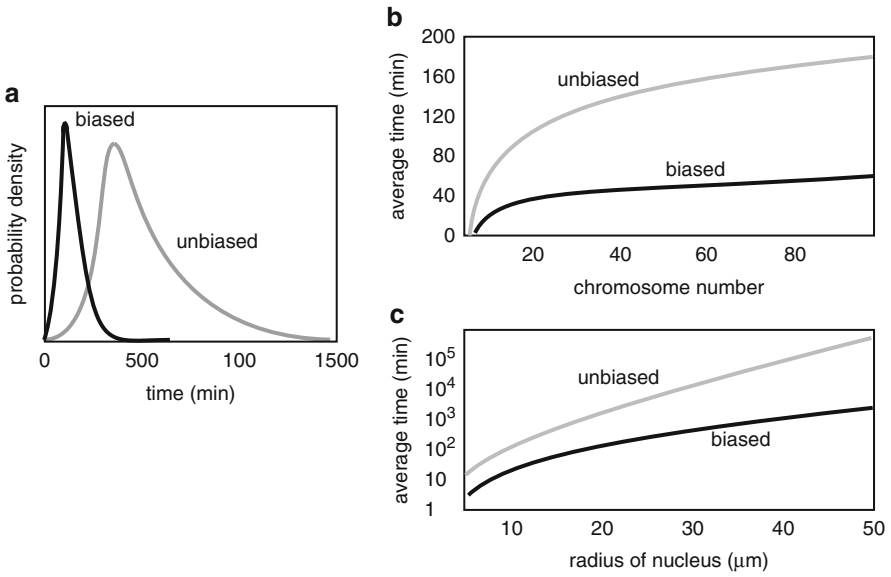


Fig. 8.9: Summary sketch of results from model simulations of Wollman et al. [685] comparing capture times for biased and unbiased search, shown by *black and gray curves*, respectively. (**a**) Distribution of capture times in the case of 250 MTs searching for 46 chromosomes. (**b**) Variation of mean capture time with the number of chromosomes for 1,000 searching MTs (unbiased search) and 250 MTs (biased search), respectively. (**c**) The unbiased model exhibits an exponential increase in capture time as a function of nucleus radius, whereas the capture time of the biased model varies as a power law (approximately cubic)

2. For each of these distances, the probability of a successful search p is calculated using equation (8.2.2).
3. The number n of unsuccessful searches is generated randomly from the geometric probability distribution $P(n) = p(1-p)^n$.
4. The duration of each unsuccessful search is generated randomly using the exponential probability distribution $P(\tau) \sim \mathrm{e}^{-t/T_{\mathrm{cycle}}}$ with T_{cycle} given by equation (8.2.3). The sum of the n random duration times is added to the successful search time x/V_g to determine the total search time for the given kinetochore.
5. The above four steps are repeated N times (for N MTs) and the smallest search time is chosen. This is then repeated for each of the kinetochores and the largest of the M search times is identified as the total capture time. (Since $N \gg M$, steric effects are ignored, that is, one neglects the fact that once an MT has found a kinetochore, the number of searching MTs is reduced.)
6. The average capture time is determined by repeating the above five steps over multiple trials.

The results obtained by Wollman et al. [685] for the above unbiased search scenario are sketched in Fig. 8.9, which shows the distribution of capture times and the variation of the mean with respect to nucleus size and number of chromosomes. Also shown are the corresponding results for a biased search, in which the catastrophe rate of MTs is modulated by a RhoGTPase concentration gradient, resulting in a significant reduction in capture time (see [685] for further details).

Finally, note that another possible mechanism for reducing the capture time would be to have some form of cooperative effect between chromosomes so that once several central chromosomes have been captured by random search, the remainder could be captured more quickly. Cooperative effects might be mediated by molecular motors and bundles of MTs nucleated from the chromosomes [445].

Box 8A. Gamma distribution.

Let $T = \sum_{j=1}^{n} \tau_j$, where the τ_j are independent, exponential random variables with mean $\bar{\tau}$, that is, τ_j has the probability density

$$p(\tau_j) = \frac{1}{\bar{\tau}} \mathrm{e}^{-\tau_j/\bar{\tau}}.$$

The probability density for the random variable T is

$$\begin{aligned} \rho(T) &= \int_0^\infty \cdots \int_0^\infty \delta(T - \sum_{l=1}^{n} \tau_l) \prod_{k=1}^{n} p(\tau_k) d\tau_k \\ &= \frac{1}{\bar{\tau}^n} \int_0^\infty \cdots \int_0^\infty \delta(T - \sum_{l=1}^{n} \tau_l) \mathrm{e}^{-\sum_{j=1}^{n} \tau_j/\bar{\tau}} \prod_{k=1}^{n} d\tau_k. \end{aligned}$$

Introduce the Fourier representation of the Dirac delta function,

$$\delta(T-\sum_{l=1}^{n}\tau_l)=\int_{-\infty}^{\infty}\mathrm{e}^{iz(T-\sum_{l=1}^{n}\tau_l)}\frac{dz}{2\pi}.$$

Substituting into the integral expression for $\rho(T)$ and reordering the multiple integral yields

$$\begin{aligned}\rho(T)&=\frac{1}{\bar{\tau}^n}\int_{-\infty}^{\infty}\mathrm{e}^{izT}\prod_{l=1}^{n}\left[\int_0^{\infty}\mathrm{e}^{-\tau_l(iz+1/\bar{\tau})}d\tau_l\right]\frac{dz}{2\pi}\\&=\frac{1}{\bar{\tau}^n}\int_{-\infty}^{\infty}\mathrm{e}^{izT}\frac{1}{(iz+\bar{\tau}^{-1})^n}\frac{dz}{2\pi}\\&=\frac{(-i)^n}{\bar{\tau}^n}\int_{-\infty}^{\infty}\mathrm{e}^{izT}\frac{1}{(z-i\bar{\tau}^{-1})^n}\frac{dz}{2\pi}.\end{aligned}$$

The remaining integral can be calculated using the calculus of residues. That is, treat z as a complex variable and close the contour in the lower-half complex plane. Recall that for any analytic function $f(z)$, the integral around a closed contour C with ω an nth order pole within the interior of C is given by

$$\frac{1}{2\pi i}\int_C f(z)\frac{dz}{(z-\omega)^n}=\frac{1}{(n-1)!}\frac{d^{n-1}}{dz^{n-1}}f(z)\bigg|_{z=\omega}.$$

Taking C to be the semicircle in the lower half-plane, $\omega=i/\bar{\tau}$, and $f(z)=\mathrm{e}^{izT}$, we see that

$$\rho(T)=\frac{1}{\bar{\tau}^n}\frac{T^{n-1}}{(n-1)!}\mathrm{e}^{-T/\bar{\tau}}.$$

8.2.2 Chromosome Movements and Directional Instability

It is found that the movements of chromosomes during prometaphase and metaphase are characterized by periods of ballistic motion or "runs" at approximately constant speed, separated by abrupt reversals in direction of movement [97, 241, 538, 596]. These oscillations arise from interactions between attached kinetochore MTs (kMTs) and the corresponding chromosomes. When a chromosome becomes *bioriented*, that is, the sister chromatids are attached to kMTs emanating from opposite poles, the duration of movements towards and away from the nearest pole is biased so that the chromosome is aligned with the spindle equator at metaphase.

Hill Sleeve Model of kMT–Kinetochore Interactions

One of the first models of kMT/chromosome interactions was developed by Hill [260], who treated the outer region of a kinetochore as a "sleeve" containing a sequence of tubulin binding sites that holds a kMT within the sleeve (see Fig. 8.10a). The sleeve is assumed to be around 40 nm thick, which means that it can accommodate up to $M = 65$ tubulin subunits of the kMT. The model keeps track of the position n of the kMT tip within the sleeve, $n = 1, \ldots, M$, with limits $n = 1$ (fully inserted) and $n = M$ (almost unattached). Motion of the kMT tip is modeled as a random walk in the free energy landscape sketched in Fig. 8.10b. Let $-a$, $a > 0$, be the binding energy of a single tubulin site. This effect tends to pull the kMT into the sleeve. However, the attractive force is opposed by two frictional forces—an external frictional force F that opposes movement of the whole chromosome and a surface roughness acting at the interface between the kMT and the kinetochore. The latter increases linearly with the length of the interface, which is proportional to $M - n + 1$. If l is the length of a single step, then the forward transition $n \to n + 1$ (withdrawal) requires climbing a barrier height of size $\Delta E_+(n) = a + (M - n + 1)b - Fl/2$, whereas the reverse transition $n + 1 \to n$ (insertion) requires climbing a barrier height of size $\Delta E_-(n) = (M - n + 1)b + Fl/2$. (For convenience, the external frictional force is divided equally between the two transitions.) From the theory of chemical kinetics (Sect. 3.1), the forward and backward transition rates take the form

$$k_\pm(n) = \kappa \mathrm{e}^{-\Delta E_\pm(n)/k_B T},$$

for a background hopping rate κ. The position of the tip can also change due to polymerization or depolymerization. We will assume that the rate of adding a tubulin subunit at the tip is αc, where c is the background concentration of tubulin, and the rate of removing a subunit is $\beta \mathrm{e}^{-a/k_B T}$. Here α, β are constants and the Boltzmann factor takes into account the binding energy of the subunit to the kinetochore sleeve. We can now write down a kinetic equation for the mean tip location $n(t)$:

$$\begin{aligned} \frac{dn}{dt} &= k_+(n) - k_-(n) + \beta \mathrm{e}^{-a/k_B T} - \alpha c \\ &= \kappa \mathrm{e}^{-[(M-n+1)b + a - Fl/2]/k_B T} + \beta \mathrm{e}^{-a/k_B T} - \kappa \mathrm{e}^{-[(M-n+1)b + Fl/2]/k_B T} - \alpha c. \end{aligned}$$

Hence, at equilibrium the mean tip location n^* is obtained by setting $\dot{n} = 0$:

$$\kappa \mathrm{e}^{-(M-n+1)b/k_B T} \left[\mathrm{e}^{-Fl/2k_B T} - \mathrm{e}^{Fl/2k_B T} \mathrm{e}^{-a/k_B T} \right] = \beta \mathrm{e}^{-a/k_B T} - \alpha c$$

which gives

$$n^* = M + 1 - \frac{k_B T}{b} \ln \Gamma, \quad \Gamma = \kappa^{-1} \frac{\beta \mathrm{e}^{-a/k_B T} - \alpha c}{\mathrm{e}^{-Fl/2k_B T} - \mathrm{e}^{Fl/2k_B T} \mathrm{e}^{-a/k_B T}}. \tag{8.2.6}$$

Suppose that depolymerization dominates over polymerization so that the tendency of the kinetochore is to move towards the nearest pole. An important feature

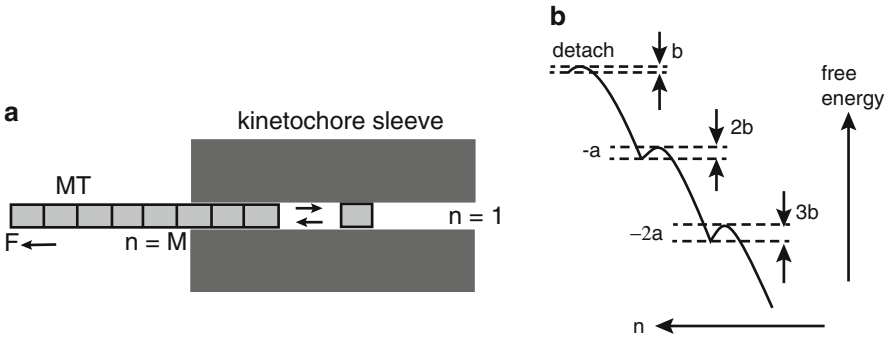


Fig. 8.10: Hill "sleeve model." (**a**) Schematic diagram of a kinetochore sleeve interacting with a kMT. The position of the kMT tip changes either due to thermal motion of the sleeve or by addition or loss of tubulin subunits at the kMT tip. (**b**) Free energy diagram for a kinetochore sleeve interacting with a kMT in the absence of external friction ($F = 0$)

of Hill's sleeve model is that over a wide range of external forces F, the speed of depolymerization-coupled kinetochore movements remains constant . This is a consequence of the observation that for a given F, the steady-state position $n^*(F)$ of the kMT tip within the sleeve is fixed, and thus the sleeve moves at an average speed equal to the rate of kMT shortening. In other words, the sleeve keeps up with the tip of the depolymerizing MT. If the external force changes, $F \to F'$, the sleeve will shift to a new steady-state position $n^*(F')$, where it will continue on at the rate of kMT shortening. This approximate load independence is consistent with experimental observations [596].

Recently, a more mathematical treatment of the Hill sleeve has been developed by Shtylla and Keener based on a jump-diffusion process [588, 589]. Suppose that the position of the kMT tip relative to the entrance of the kinetochore is a continuous random variable $X(t)$. The SDE for $X(t)$ is taken to be

$$dX(t) = \frac{1}{\gamma}(-\Psi'(X) - F)dt + \sqrt{2D}dW(t) + \delta\, dN_\alpha(t) + \delta\, dN_\beta(t), \qquad (8.2.7)$$

where $W(t)$ is a standard Wiener process, D is the diffusion coefficient, and γ is the drag coefficient satisfying the Einstein relation $D = k_B T/\gamma$. Here $N_\alpha(t)$ and $N_\beta(t)$ are independent homogeneous Poisson processes with amplitudes δ and position-dependent rates $\alpha(x)$ and $\beta(x)$, respectively, which model the addition and removal of tubulin monomers of size δ from the kMT tip; this subsequently shifts the position of the tip by integer multiples of δ. Finally, $\Psi'(x)$ represents the binding force due to interactions between the kMT and kinetochore with $\Psi(x)$ the corresponding free energy (see Fig. 8.10b), and F denotes any loads on the kMT. For concreteness, $\Psi(x)$ is taken to have the corrugated form [589]

$$\Psi(x) = f(x)(1 - \cos(2\pi x/\delta)) + h(x),$$

where

$$f(x) = \frac{bx}{2\delta} + C, \quad h(x) = -ax/\delta, \text{ for } 0 < x \leq L_1$$

and

$$f(x) = f(L_1), \quad h(x) = h(L_1) \text{ for } L_1 < x \leq L,$$

where L_1 is the length of the sleeve region along which binding sites occur, C is a constant, and a, b are as depicted in Fig. 8.10b. Note that the spacing of binding sites is taken to be equal to the monomer size δ. It is also possible to generalize the model so that these two quantities are incommensurate [589]. The differential Chapman–Kolmogorov (CK) equation for the given jump-diffusion process is

$$\begin{aligned}\frac{\partial p(x,t)}{\partial t} = &-\frac{1}{\gamma}\frac{\partial}{\partial x}\left(V'(x)p(x,t)\right) + D\frac{\partial^2}{\partial x^2}p(x,t) + \alpha(x-\delta)p(x-\delta,t) \\ &+ \beta(x+\delta)p(x+\delta,t) - (\alpha(x)+\beta(x))p(x,t),\end{aligned} \tag{8.2.8}$$

where $V(x) = -\Psi(x) - Fx$. Equation (8.2.8) is supplemented by a reflecting boundary condition at $x = L$, where L is the length of the sleeve, and an absorbing boundary condition at $x = 0$, since if the kMT crosses this point, it detaches from the kinetochore and is unlikely to reattach. The attachment and detachment rates are taken to be steep sigmoid functions of position:

$$\alpha(x) = \frac{\alpha_0}{1 + e^{\lambda(x-\alpha_1)}}, \quad \beta(x) = \beta_0 + \frac{\beta_2 - \beta_0}{1 + e^{-\lambda(x-\beta_1)}},$$

with $\lambda \gg 1/\delta$ and $x_0 \approx L - \gamma$. Thus $\alpha(x) \approx \alpha_0$ for $x < \alpha_1$ and rapidly drops to zero as it reaches the end of the sleeve at $x = L$ (see Fig. 8.11). On the other hand $\beta(x)$ switches from a lower value β_0 to a higher value β_2 at an intermediate location $x = \beta_1$. This could be due to the presence of a depolymerase in the region proximal to the end of the sleeve. In the absence of other factors, the system would reach an equilibrium position x_0 where the rates of polymerization and depolymerization balance, that is, $\alpha(x_0) = \beta(x_0)$. If $\beta_2 > \alpha_0$ (as depicted in Fig. 8.11), then $x_0 \approx \beta_1$ and the system is called a depolymerization Kt motor, whereas if $\beta_2 < \alpha_0$, then $x_0 \approx \alpha_1$ and we have a polymerization Kt motor.

Monte Carlo simulations of the stochastic model (8.2.7) indicate that the mean time for the kMT to detach from the kinetochore when starting at $x = L$ is much larger than the relaxation time of the system, provided that the load force F is not too large [589]. That is, given biophysically reasonable parameter values, one finds the MFPT is $\langle T(L)\rangle \sim 100s$, whereas the relaxation time is $\tau \sim 1s$. Therefore, it makes sense to determine a (quasi) steady-state probability density $p_s(x)$ by replacing the absorbing boundary at $x = 0$ by a reflecting boundary. Note that the resulting steady-state CK equation can be solved analytically using multi-scale analysis [589] (see Box 8B). One finds that there are qualitative differences in the properties of $p_s(x)$ as the parameter $k = b/a$ is varied. For relatively large k ($k \approx 0.1$), polymerization/depolymerization dominates diffusion due to the large barrier height that must be overcome to shift along the sleeve. Thus one finds $p_s(x)$ is localized around the point $x = x_0$, and the distribution is relatively insensitive to the load force F.

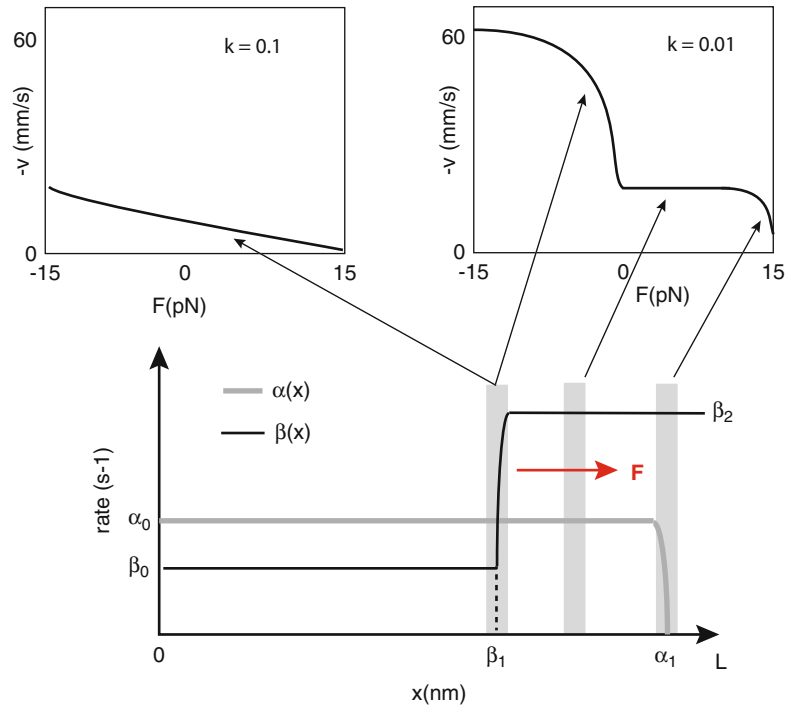


Fig. 8.11: Continuum sleeve model [589]. Sketch of space-dependent rate functions $\alpha(x)$ (polymerization) and $\beta(x)$ (depolymerization) of the kMT tip. *Vertical shaded* regions indicate where the stationary density $p_s(x)$ is localized. If $k = b/a$ is sufficiently small (inset with $k = 0.01$), then the stationary density $p_s(x)$ shifts in response to change in the load force F (as indicated by *red arrow*). If $p_s(x)$ is localized in a region where $\alpha(x) - \beta(x)$ is constant, then the velocity is independent of F whereas the velocity changes rapidly with F in regions where $\alpha(x) - \beta(x)$ changes rapidly. For larger values of k (inset with $k = 0.1$), one finds that the density $p_s(x)$ only changes slightly as a function of F and is localized around $x = x_0$ where $\beta(x_0) = \alpha(x_0)$. Since this occurs in a region where $\alpha(x) - \beta(x)$ changes rapidly, the velocity is sensitive to the value of F

On the other hand, for smaller values of k, diffusion plays an important role and the stationary distribution tends to be localized around an F-dependent point that minimizes the total free energy. Positive (negative) F shifts $p_s(x)$ towards $x = 0$ ($x = L$). Given the stationary density $p_s(x)$, the mean velocity of the kMT/kinetochore system relative to an outside frame of reference is

$$v = \delta \int (\alpha(y) - \beta(y)) p_s(y) dy. \tag{8.2.9}$$

Since the kinetochore and kMT interactions have reached steady state, net depolymerization (polymerization) of kMT means that the Kt motor system moves towards (away from) the corresponding spindle pole, which corresponds to a negative (positive) mean velocity. One also finds qualitative differences in the load–velocity

relationship as k varies. For small k, changes in F shift the probability density $p_s(x)$. However, if this occurs in regions where $\alpha(x)-\beta(x)$ is x-independent, then the velocity does not change. Thus one finds significant ranges of F over which v does not change, which again is consistent with experimental observations [596]. On the other hand, for larger values of k the load–velocity relationship becomes more monotone. In this "sticky" motor regime attachment can be maintained against large loads but at the cost of reduced velocities. The qualitative behavior is illustrated in Fig. 8.11 for a depolymerizing Kt motor. Similar results hold for a polymerizing motor, except that one typically finds positive velocities for positive F, since $p_s(x)$ is shifted to a region where polymerization dominates.

Box 8B. Asymptotic analysis of the continuum Hill model in cell mitosis.

We derive an approximate steady-state solution of the CK equation (8.2.8) following the multi-scale analysis presented in [589]. The steady-state CK equation for $p(x)$ is

$$0 = -\frac{1}{\gamma}\frac{\partial}{\partial x}\left(V'(x)p(x)\right) + D\frac{\partial^2}{\partial x^2}p(x) + \alpha(x-\delta)p(x-\delta) + \beta(x+\delta)p(x+\delta) - (\alpha(x)+\beta(x))p(x), \tag{8.2.10}$$

which is supplemented by reflecting boundaries at $x = 0, L$. Setting

$$V(x) = f(x)\cos(2\pi x/\delta) + r(x), \quad r(x) = -f(x) - h(x) - Fx,$$

Taylor expanding the jump terms as infinite series in δ and then integrating with respect to x using the reflecting boundary conditions give

$$0 = -\frac{1}{\gamma}\frac{d}{dx}\left(f(x)\cos(2\pi x/\delta) + r(x)\right)p(x) + D\frac{d}{dx}p(x) + \sum_{n=1}^{\infty}\frac{\delta^n}{n!}\frac{d^{n-1}}{dx^{n-1}}\left([\beta(x) + (-1)^n\alpha(x)]p(x)\right).$$

Nondimensionalize the CK equation by setting

$$x = \eta y, \quad \eta = \frac{\gamma D\delta}{b}$$

with $D\gamma = k_B T$ and introduce the small parameter $\varepsilon = b/k_B T$. This gives

$$0 = -\frac{d}{dy}\left(\hat{f}(y)\cos(2\pi y/\varepsilon) + \hat{r}(y)\right)\hat{p}(y) + \frac{d}{dy}\hat{p}(y) + \alpha_2 \sum_{n=1}^{\infty} \frac{\varepsilon^{n-1}}{n!}\frac{d^{n-1}}{dy^{n-1}}\left([\hat{\beta}(y) + (-1)^n\hat{\alpha}(y)]\hat{p}(y)\right), \tag{8.2.11}$$

with $\alpha_2 = \beta_0\delta^2\gamma/b$ and

$$\hat{p}(y) = p(x),\ \hat{f}(y) = f(x)/k_BT,\ \hat{r}(y) = r(x)/k_BT,$$
$$\hat{\beta}(y) = \beta(x)/\beta_0,\ \hat{\alpha}(y) = \alpha(x)/\beta_0.$$

In the following we drop the ˆon the functions $f, r, \alpha, \beta, and p$.

It can be seen from equation (8.2.11) that the drift term consists of a high spatial frequency periodic oscillator with a slowly varying amplitude. This suggests using the method of multiple scales. That is, we introduce a slow spatial variable $z = y$ and a fast spatial variable $\sigma = y/\varepsilon$, which are treated as independent variables. That is, $p(y) = p(z,\sigma)$ with

$$\frac{d}{dy} = \frac{\partial}{\partial z} + \frac{1}{\varepsilon}\frac{\partial}{\partial \sigma}.$$

Equation (8.2.11) becomes

$$0 = -\left(f'(z)\cos(2\pi\sigma) + r'(z) - \frac{2\pi}{\varepsilon}f(z)\sin(2\pi\sigma)\right)p(z,\sigma) + \frac{\partial}{\partial z}p(z,\sigma) + \frac{1}{\varepsilon}\frac{\partial}{\partial\sigma}p(z,\sigma) + \alpha_2\sum_{n=1}^{\infty}\frac{1}{n!}[\beta(z) + (-1)^n\alpha(z)]\frac{\partial^{n-1}}{\partial\sigma^{n-1}}p(z,\sigma), \tag{8.2.12}$$
$$+ \varepsilon\alpha_2\sum_{n=1}^{\infty}\frac{n-1}{n!}\frac{\partial}{\partial z}\frac{\partial^{n-1}}{\partial\sigma^{n-1}}([\beta(z) + (-1)^n\alpha(z)]p(z,\sigma)) + O(\varepsilon^2).$$

We now seek an asymptotic solution of the form

$$p = p_0(z,\sigma) + \varepsilon p_1(z,\sigma) + O(\varepsilon^2),$$

with $p_j(z,\sigma)$ a one-periodic function of σ. Substituting the asymptotic expansion into equation (8.2.12) and collecting same order terms in ε generates a hierarchy of equations. The $O(1/\varepsilon)$ equation is

$$2\pi f(z)\sin(2\pi\sigma)p_0(z,\sigma) + \frac{\partial}{\partial\sigma}p_0(z,\sigma) = 0, \tag{8.2.13}$$

and the $O(1)$ equation is

$$
\begin{aligned}
&2\pi f(z)\sin(2\pi\sigma)p_1(z,\sigma)+\frac{\partial}{\partial\sigma}p_1(z,\sigma)\\
&=\left(f'(z)\cos(2\pi\sigma)+r'(z)\right)p_0(z,\sigma)-\frac{\partial}{\partial z}p_0(z,\sigma)\\
&\quad-\alpha_2\sum_{n=1}^{\infty}\frac{1}{n!}[\beta(z)+(-1)^n\alpha(z)]\frac{\partial^{n-1}}{\partial\sigma^{n-1}}p_0(z,\sigma).
\end{aligned}
\tag{8.2.14}
$$

These are supplemented by the $O(\varepsilon)$ normalization condition

$$
\int_0^{L/\eta}\left(p_0(z,\sigma)+\varepsilon p_1(z,\sigma)\right)dz=1. \tag{8.2.15}
$$

Equation (8.2.13) may be solved by direct integration to give

$$
p_0(z,\sigma)=A_0(z)\exp\left(f(z)\cos(2\pi\sigma)\right). \tag{8.2.16}
$$

The slowly varying amplitude $A_0(z)$ is then determined from the $O(1)$ equation (8.2.14). It is first convenient to sum the infinite series in (8.2.14) using the following identities for the function $F(\sigma)=\int_0^{\sigma}p_0(z,s)ds$ with z fixed:

$$
\begin{aligned}
\int_0^1 p_0(z,s)ds&=F(\sigma+1)-F(\sigma)=\sum_{n=1}^{\infty}\frac{1}{n!}\frac{\partial^n}{\partial\sigma^n}F(\sigma)\\
&=\sum_{n=1}^{\infty}\frac{1}{n!}\frac{\partial^{n-1}}{\partial\sigma^{n-1}}p_0(z,\sigma)
\end{aligned}
$$

and

$$
\begin{aligned}
-\int_0^1 p_0(z,s)ds&=F(\sigma-1)-F(\sigma)=\sum_{n=1}^{\infty}\frac{(-1)^n}{n!}\frac{\partial^n}{\partial\sigma^n}F(\sigma)\\
&=\sum_{n=1}^{\infty}\frac{(-1)^n}{n!}\frac{\partial^{n-1}}{\partial\sigma^{n-1}}p_0(z,\sigma).
\end{aligned}
$$

It follows that (8.2.14) simplifies as

$$
\begin{aligned}
2\pi f(z)\sin(2\pi\sigma)p_1(z,\sigma)+\frac{\partial}{\partial\sigma}p_1(z,\sigma)&=\left(f'(z)\cos(2\pi\sigma)+r'(z)\right)p_0(z,\sigma)\\
-\frac{\partial}{\partial z}p_0(z,\sigma)-\alpha_2[\beta(z)-\alpha(z)]\int_0^1 p_0(z,\sigma)d\sigma.&
\end{aligned}
\tag{8.2.17}
$$

Multiply both sides of this equation by the integrating factor

$$I(z,\sigma) = \mathrm{e}^{-f(z)\cos(2\pi\sigma)}$$

and integrate with respect to σ. Exploiting the fact that $p_1(z,\sigma)$ is periodic in σ, so that

$$\int_0^1 \frac{\partial}{\partial\sigma}[p_1(z,\sigma)I(z,\sigma)]d\sigma = 0,$$

we obtain the following solvability condition:

$$\int_0^1 \left[\mathrm{e}^{r(z)}\frac{\partial}{\partial z}\left(p_0(z,\sigma)\mathrm{e}^{-r(z)}I(z,\sigma)\right) + \alpha_2[\beta(z)-\alpha(z)]I(z,\sigma)\int_0^1 p_0(z,s)ds\right]d\sigma = 0.$$

Substituting the explicit solution for $p_0(z,\sigma)$ we obtain an equation for the amplitude $A_0(z)$:

$$-r'(z)A_0(z) + A_0'(z) + \alpha_2 A_0(z)[\beta(z)-\alpha(z)]I_0^2(f(z)) = 0, \quad (8.2.18)$$

where $I_0(x)$ is the modified Bessel function of the first kind, whose integral representation is

$$I_0(f(z)) = \int_0^1 \mathrm{e}^{\pm f(z)\cos(2\pi\sigma)}d\sigma.$$

It follows that

$$A_0(z) = \hat{C}\exp\left(r(z) - \alpha_2\int^z I_0^2(f(z')[\beta(z')-\alpha(z')]dz'\right), \quad (8.2.19)$$

so that in terms of the original variables, the lowest order approximation of the steady-state density is

$$p_0(x) = C\exp\left(\frac{V(x)}{k_BT} - \frac{\delta\gamma}{k_BT}\int^x I_0^2(f(x')/k_BT)[\beta(x')-\alpha(x')]dx'\right) + O(\varepsilon). \quad (8.2.20)$$

The constant C can be determined using the normalization condition (8.2.15). Given $p_0(x)$, we one can numerically calculate the velocity using the integral expression

$$v = \delta\int_0^L(\alpha(x)-\beta(x))p_0(x)dx + O(\varepsilon).$$

An alternative expression for v, which is easier to evaluate numerically, is derived in Ex. 8.4. One final observation is that $p_0(x)$ is equivalent to the steady-state solution of an FP equation whose drift terms is given by the derivative of the function in the exponential. This implies that the MFPT for the kMT to detach from the sleeve can be estimated by solving the FPT problem of the reduced FP equation.

Chromosomal Oscillations

As it stands, the sleeve model of single kMT–kinetochore interactions cannot explain the oscillatory switching of chromosome motion towards and away from a spindle pole (see Fig. 8.12). In order to account for such dynamics, it is necessary to consider some combination of the following mechanisms: tensional coupling between sister chromatids, polar ejection forces due to the pushing action of polar (non-kinetochore) MTs on the arms of a chromosome, the space-dependent modulation of kMT catastrophe and rescue frequencies, and the possible action of depolymerizing and polymerizing molecular motors such as kinesin and dynein [120, 205, 207, 303, 394, 588]. Joglekar and Hunt [303] considered a simple extension of the Hill sleeve model by including tensional coupling and polar ejection forces and allowing detached MTs to switch from catastrophe to rescue. Following the Hill model, kMTs are assumed to be in catastrophe so that in order to maintain an average kMT tip location within the kinetochore, the associated chromatid tends to move towards the pole of the attached kMT. It follows that there is essentially a ToW between the sister chromatids moving in opposite directions. As the sister chromatids start to separate, tensional forces increase until one of the chromatids loses all of its depolymerizing kMTs, at which point it follows the motion of the other chromatid. Consequently the chromosome moves towards the pole of the winning chromatid. However, as it approaches the pole, the density of polar MTs increases, resulting in an increase in the polar ejection force. The closer the kinetochore moves towards the pole, the more strongly the polar ejection force opposes its advancement, and eventually the last depolymerizing kMT detaches from the kinetochore as the load exceeds the detachment force. Meanwhile growing MTs from the opposite pole can be recruited by the sister chromatid resulting in an abrupt reversal in direction and the cycle repeats. An immediate consequence of the interplay between polar ejection forces, MT dynamic instabilities, and the "Hill sleeve" is that the chromosomes tend to position themselves towards the spindle equator.

One of the limitations of the Joglekar and Hunt model is that it neglects the experimental observation that there is a tension-dependent increase in the rescue frequency of kMTs. This particular mechanism has been explored in a number of computational models by Gardner et al. [205, 207, 606]. A schematic illustration of the basic model is shown in Fig. 8.13; only one chromatid pair is shown for simplicity. The minus end of each kMT is fixed at its spindle pole and the plus end of the kMT

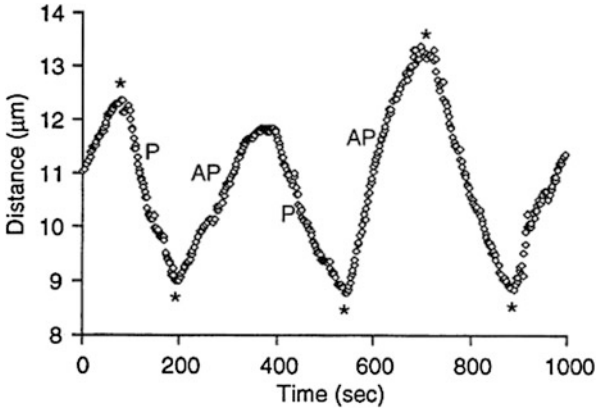


Fig. 8.12: Sawtooth-like in vivo oscillations of a non-oriented chromosome (attached to one spindle pole) observed in a newt mitotic cell. The chromosome switches between poleward (P) and anti-poleward (AP) motion (Adapted from Inoue et al. [291])

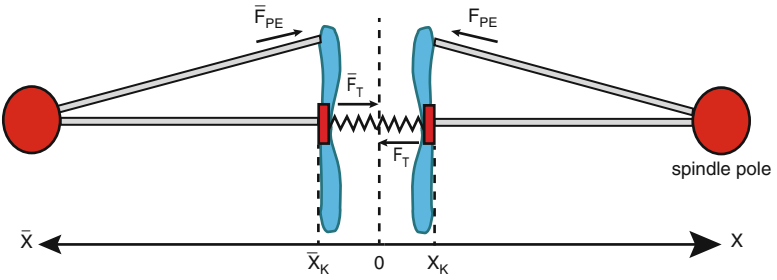


Fig. 8.13: Tension-mediated regulation of kMT/kinetochore dynamics. Here $F_{\rm T}$ and $F_{\rm PE}$ denote the tension force and polar ejection force acting on the rightward kMT whose tip is at position X_K relative to the equator. Similarly, $\overline{F}_{\rm T}$ and $\overline{F}_{\rm PE}$ denote the corresponding forces on the leftward kMT whose tip is at $\overline{X}_K$. Note the sign conventions

is fixed at the corresponding kinetochore; thus the details of the kMT–kinetochore interface, as specified in the Hill sleeve model, are neglected. For convenience, distance from the spindle equator towards the right (left) pole is denoted by the positive coordinate x ($\bar{x}$), and the direction of all forces F ($\overline{F}$) acting on the right-half (left-half) of the spindle are defined with respect to the poleward direction. Let $X_{\rm K}$ and $\overline{X}_{\rm K}$ denote the current position of the right and left sister kinetochore plates with respect to the spindle equator, and let $V_{\rm K}$ and $\overline{V}_{\rm K}$ denote the corresponding velocities. For concreteness, let us focus on a kMT attached to the right-hand pole; a similar formulation holds for those attached to the other pole. Each kMT is modeled in terms of the Dogterom–Leibler model of microtubule catastrophe and rescue (see equation (4.1.10) and [146]), so that the mean velocity is

$$\langle V_K \rangle = \frac{k_- v_+ - k_+ v_-}{k_+ + k_-},$$

where k_+ (k_-) is the catastrophe (rescue) frequency and v_+ (v_-) is the rate of growth (shrinkage). The catastrophe and rescue frequencies $k_\pm$ are assumed to be regulated by mechanical forces acting on the kMT–kinetochore system, perhaps in combination with a chemical gradient in some kMT catastrophe promoter, and are thus x-dependent. The basic components of the model are as follows [606]:

1. *Tension force between the kinetochores* $F_{\rm T}$. The tension force is modeled as a linear spring, so that the magnitude of the force depends on the distance between the sister kinetochores according to

$$F_{\rm T} = \kappa(X_{\rm K} + \overline{X}_{\rm K} - d_0), \tag{8.2.21}$$

   where κ is the spring constant and d_0 is the equilibrium length of the spring. Note that, given our sign convention, $F_{\rm T} = \overline{F}_{\rm T}$. The tension force provides the coupling between the two kinetochores.

2. *Polar ejection force* $F_{\rm PE}$. This is directed towards the spindle equator and is due to the interaction between the chromosome arms and the plus ends of polar MTs. Hence, it is proportional to the density of polar MTs emanating from the pole. From geometric arguments one can take

$$F_{\rm PE} = \rho X_{\rm K}^2 \tag{8.2.22}$$

   for some constant ρ.

3. *Rescue frequency.* The rescue frequency is taken to be a function of the total force $F = F_{\rm T} + F_{\rm PE}$ according to

$$k_- = k_{-,0} \mathrm{e}^{F/k_B T}.$$

4. *Chemical gradient.* It is assumed that a kMT catastrophe promoter forms a spatial gradient due to spatial segregation of a kinase/phosphatase system that regulates the promoter; the kinase phosphorylates (deactivates) the promoter, whereas the phosphatase dephosphorylates (activates) the promoter. The spatial gradient is modeled by taking the kinase to be localized to the surface of the spindle poles, whereas phosphatase is distributed homogeneously throughout the cell volume. Let c_A and c_B denote the concentration of deactivated and activated catastrophe promoter, respectively, such that $c_A(x,t) + c_B(x,t) = c_T$ with c_T a constant. The reaction–diffusion equation for c_A is

$$\frac{\partial c_A}{\partial t} = \frac{\partial^2 c_A}{\partial x^2} - k c_A,$$

   supplemented by the boundary conditions

$$-D\frac{\partial c_A}{\partial x}\bigg|_{x=0} = k^* c_B(0), \quad -D\frac{\partial c_A}{\partial x}\bigg|_{x=L/2} = 0.$$

Here k and k^* are the rates of activation and deactivation due to the action of kinds/phophotase, and L is the spatial separation of the poles. The analysis of intracellular spatial gradients is discussed more fully in Sect. 9.1.1. One finds that the steady-state concentration is of the form

$$c_A(x) = c_T \left(A e^{-\gamma x/L} + B e^{\gamma x/L} \right), \quad \gamma = 2\sqrt{k/D}.$$

5. *Catastrophe frequency.* Given the steady-state spatial gradient $c_B(x)$ of catastrophe promoter, the catastrophe frequency of the kMT tip is taken to be

$$k_+ = k_{+,0} + \beta c_B(X_K).$$

Computer simulations of the above model, involving different combinations of the various components, suggest that in order to match experimental data of chromosomal dynamics during metaphase of budding yeast, including the tendency to cluster at the cell equator, it is necessary to combine MT dynamic instabilities with tension-dependent rescue and either polar ejection forces or a spatial gradient in some catastrophe promoter [207, 606]. Finally, note that some more recent theoretical studies have considered explicit models of the mechanobiochemical feedback mechanism, in which tension-dependent molecular sensor molecules enzymatically regulate a phosphorylation cascade that alters the dephosphorylation rate at the kMT tip [394, 588]. Other recent models have emphasized the role of chromokinesin motors on polar MTs that exert a polar ejection force on the chromosomal arms [94, 617].

8.2.3 Force Balance and Spindle Length Control

It is hypothesized that the interplay between kMT dynamic instability, kMT–kinetochore interactions, various populations of molecular motors, and elastic or viscoelastic forces plays an important role in later stages of mitosis. This includes the maintenance of spindle length (distance between the poles) in metaphase, the separation of sister chromatids in anaphase A, and the increase in spindle length in anaphase B [119, 232, 428]. For example, during anaphase A, there is a net poleward flux of kMTs due to depolymerization at their pole-associated minus ends, which may be supplemented by a so-called "pacman" mechanism in which the kinetochores actively "chew" their way towards the poles by depolymerization of kMTs at their plus ends. On the other hand, during anaphase B, a new subset of motors known as ipMTs drive spindle elongation. Examples of changes in spindle length during different stages of mitosis are shown in Fig. 8.14 for several organisms. A number of computational models have been developed in order to study one or more of these stages by considering the various forces acting on kMTs and the chromosomes [84, 120, 131, 182, 233]. Here we will illustrate the flavor of force-

balance models by considering in more detail the model of chromosome motility in *Drosophila* embryos [120]. Figure 8.15 is a schematic diagram indicating the various forces acting on the kinetochores and kMTs of the mitotic spindle. For simplicity, we only consider one kMT per kinetochore, although it is possible to extend the model to multiple kMTs [120]. The sign convention of spatial coordinates and forces are as in Fig. 8.13. In contrast to the models of Gardner et al. [207, 606], the tip of a kMT can move relative to its kinetochore, so that one must now consider separate force-balance equations for the velocities of the tips and kinetochore. Moreover, polar ejection and tension forces are supplemented by several additional forces including the action of kinesin and dynein motors at the kMT tip.

Following the formulation of the Gardner et al. models, let $X_{\rm K}$ and $\overline{X}_{\rm K}$ denote the current position of the right and left sister kinetochore plates with respect to the spindle equator, and let $V_{\rm K}$ and $\overline{V}_{\rm K}$ denote the corresponding velocities. Similarly, let $X_{\rm MT}$ and $\overline{X}_{\rm MT}$ be the current position of the plus ends of the right and left kMTs with respect to the spindle equator, and denote the corresponding poleward sliding rates by $V_{\rm MT}$ and $\overline{V}_{\rm MT}$. (Note that $V_K = \dot{X}_K$, whereas the relationship between $X_{\rm MT}$ and $V_{\rm MT}$ is more complicated due to dynamic instability at the plus end of a kMT; see below.) The force-balance equations for the right and left kinetochores (in a low Reynolds number regime) are given by

$$\mu V_{\rm K} = F_{\rm K} - F_{\rm P} - F_{\rm T} - F_{\rm PE} \tag{8.2.23a}$$

$$\mu \overline{V}_{\rm K} = \overline{F}_{\rm K} - \overline{F}_{\rm P} - \overline{F}_{\rm T} - \overline{F}_{\rm PE} \tag{8.2.23b}$$

where μ is the drag coefficient and $F_{\rm T}$ and $F_{\rm PE}$ are given by equations (8.2.21) and (8.2.22). The additional forces on the right kinetochore are as follows (with analogous definitions for the left kinetochore):

1. *Net kinetochore motor force* $F_{\rm K}$. Let $n_\pm$ denote the density of bound plus-end (minus-end) moving kinetochore motors. These motors act on the region of the MT tip that is inserted into the kinetochore; this region has length $R - [X_{\rm MT} - X_{\rm K}]$, where R is the length of the kinetochore sleeve. If $f_\pm$ denotes the force generated by each plus-and minus-end-directed motor, then

$$F_{\rm K} = (X_{\rm K} + R - X_{\rm MT})(n_- f_- - n_+ f_+). \tag{8.2.24}$$

The motors are assumed to obey linear force–velocity relations (see Sect. 4.4.2),

$$f_\pm = F_\pm \left(1 - \frac{v_\pm}{V_\pm}\right), \tag{8.2.25}$$

where $F_\pm$ is the stall force, $V_\pm$ is the motor velocity without a load, and $v_\pm$ is the current velocity of the motors with

$$v_- = -v_+ = V_{\rm K} - V_{\rm MT}. \tag{8.2.26}$$

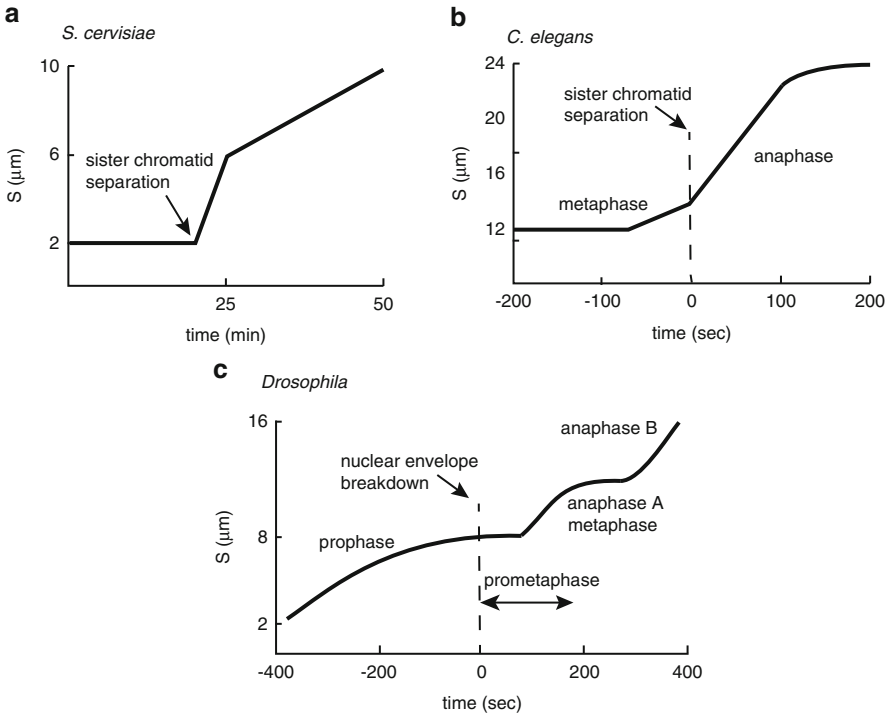


Fig. 8.14: Changes in spindle length $S(t)$ with time t during formation of mitotic spindles for three cell types: (**a**) budding yeast (Saccharomyces cerevisiae) from metaphase through anaphase B—the latter is biphasic; (**b**) Caenorhabditis elegans single-cell embryos; and (**c**) Drosophila melanogaster embryos. Redrawn from Goshima and Scholey [232]

2. *Plus-end polymerization force* $F_{\rm P}$. This will only arise if the MT impinges on the kinetochore plate, that is, $X_{\rm K} > X_{\rm MT}$, in which case it is taken to have the linear form

$$F_{\rm P} = \varepsilon(X_{\rm K} - X_{\rm MT}), \tag{8.2.27}$$

where ε is the elastic modulus of the plate.

In order to obtain a closed set of equations, it is necessary to determine the kMT sliding velocity $V_{\rm MT}$ and the position of the tip $X_{\rm MT}$. The former is determined by considering the force-balance equation for the kMTs. It turns out that the viscous drag on the kMT is negligible compared to other forces. The other forces acting on a kMT are the counter motor force $-F_{\rm K}$, the counter polymerization force $-F_{\rm P}$, and a depolymerization force $F_{\rm DP}$ at the minus end. The latter is assumed to be generated by a set of depolymerization motors with a linear force–velocity relationship. At steady-state the depolymerization velocity at the minus end of the kMT due to the action of the motors is equal to the sliding velocity $V_{\rm MT}$. It follows that

$$n_0 F_0(1 - V_{\rm MT}/V_0) = F_{\rm K} - F_{\rm P}, \tag{8.2.28}$$

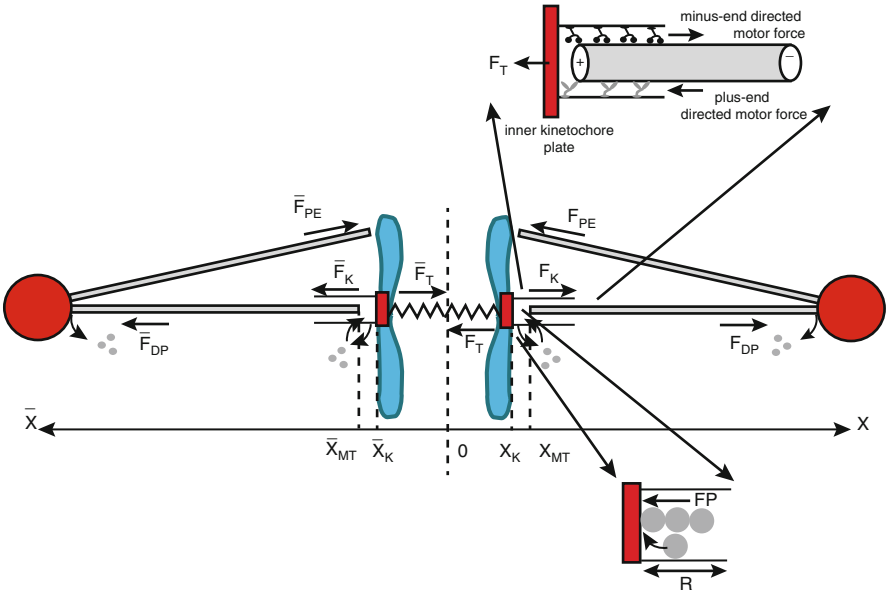


Fig. 8.15: Force balance model. See text for details. Redrawn from [120]

where n_0 is the number of active depolymerization motors, F_0 is the stall force, and V_0 is their maximum velocity. The position X_{MT} will spend both on the sliding velocity and the removal or addition of tubulin subunits at the tip. As in the Gardner et al. model [207, 606], it is assumed that the plus end undergoes dynamic instability, characterized by the stochastic switching of microtubules between growing and shrinking phases, that is between rescue and catastrophe see (Sect. 4.1). The dynamic stability is specified by four parameters given by the growth and shrinkage velocities v_g, v_s and the rates of rescue and catastrophe k_{cat} and k_{res}. The standard catastrophe model of MT dynamics is supplemented by the action of depolymerase enzymes. When tension on the kMT is low, the depolymerase acts freely on the MT plus end by suppressing the rescue frequency by some factor $\gamma > 1$. On the other hand, when the tension is high, the action of the depolymerase is blocked by some form of structural change of the kMT so that the rescue frequency recovers proportionally to the tension force. Finally, if the MT plus end is in contact with the kinetochore plate, then the catastrophe frequency is scaled up, whereas the rescue frequency returns to its low tension value.

Extensive numerical simulations of the above model have shown that it provides a quantitative description of the rapid, highly dynamic properties of metaphase and anaphase A that have been observed experimentally during *drosophila* embryo mitosis [120] (see Fig. 8.16). It also provides a possible scenario for the switch from metaphase to anaphase A based on the degradation of the cohesive bonds between the sister kinetochores and the removal of PE forces. One prediction of the model

is that increasing the level of dynein activity suppresses chromosome oscillations in metaphase. The basic idea is that the increased minus-end-directed force pulls kMTs further into the kinetochores, which increases the distance between sister chromatids and thus increases the tension on the kMTs. This promotes kMT rescue and stabilizes the kMTs within the kinetochores—recall that the basic mechanism for chromosome oscillations is depolymerization of kMT tips within the kinetochore. One limitation of the model is that it assumes that the spindle poles are fixed and thus does not provide an explanation of spindle length maintenance during metaphase. Another limitation is that it neglects the possible role of external factors such as morphogen gradients and the dynamics of a postulated viscoelastic spindle matrix.

In order to develop a model of spindle length control, it is necessary to consider the balance of forces acting on the spindle poles. Here we describe a simple force-balance model due to Goshima et al. [232, 233] (see Fig. 8.17). Let $S(t)$ and $L(t)$ denote the spindle length and length of the overlapping region of interpolar Mts (ipMTs) where kinesin-5 motors act to separate the poles. The main components of the model are the net forces acting on the poles and the polymerization/depolymerization of ipMTs. Let V_{poly} be the rate of polymerization of the ipMT plus ends, V_{depoly} be the rate of depolymerization of all MTs at the centrosome, and V_{slide} the rate at which antiparallel ipMTs slide apart. It follows that

$$\frac{dS}{dt} = 2(V_{\text{slide}} - V_{\text{depol}}), \tag{8.2.29}$$

and

$$\frac{dL}{dt} = 2(V_{\text{poly}} - V_{\text{slide}}). \tag{8.2.30}$$

The factors of two reflect the fact that pairs of ipMTs overlap. These equations are coupled to a force-balance equation (see Fig. 8.17),

$$\mu \frac{dS}{dt} = 2\left(F_{\text{slide}} - F_{\text{Kt}} - F_{\text{E}}\right). \tag{8.2.31}$$

Here F_{slide} is the kinesin-5-dependent force that slides apart the ipMTs and is given by the linear force-velocity relationship

$$F_{\text{slide}} = \alpha L\left(1 - \frac{V_{\text{slide}}}{V_{\text{max}}}\right),$$

where α is the concentration of motors. The elastic restoring force F_{E} is assumed to behave like a Hookean spring with

$$F_{\text{E}} = \beta(S - S_0),$$

where β is the effective spring constant. Finally, F_{Kt} is taken to be a constant kinetochore force that tends to pull the kinetochore and pole together via some Hill sleeve mechanism, for example. As it stands, such a model does not support a stable steady-

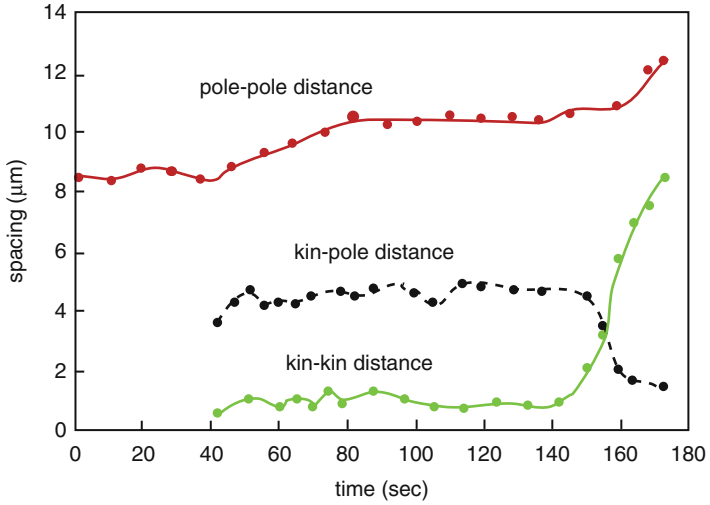


Fig. 8.16: Dynamics of spindle poles and chromatids in *Drosophila* embryos. During metaphase (80–135 s), the chromatids remain at the spindle equator and do not exhibit oscillations between the spindle poles as observed in some other organisms. During anaphase A (135–175 s), chromatids rapidly move towards the spindle poles, which are held at an approximately constant spacing of 10 μm. Redrawn from Civelekoglu-Scholey et al. [120]

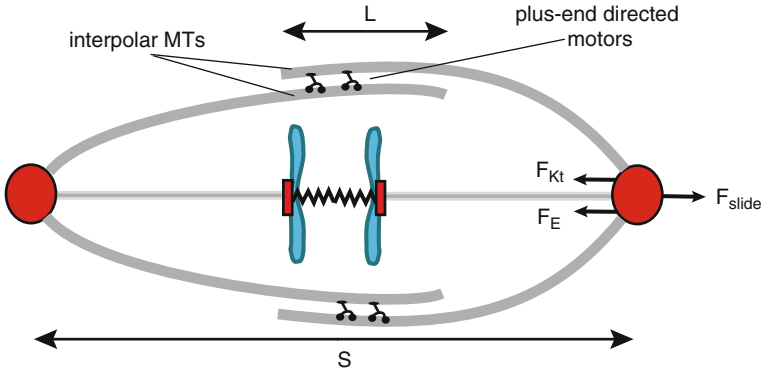


Fig. 8.17: Schematic diagram of force-balance model of spindle length control. Kinesin-5 motors slide apart antiparallel inter polar Mts (ipMTs) generating a net sliding force F_{slide} on a spindle pole that acts to separate the centrosomes. This is counteracted by a kinetochore reaction force F_{Kt} that tends to pull the kinetochore and pole together due to a Hill sleeve mechanism, for example, and a hypothesized elastic restoring force F_{E}, which could be due to MT elasticity or a spindle matrix. The spatial separation of the poles is denoted by S, whereas the length of ipMT overlap is given by L

state length S, but can be used to model spindle length separation during anaphase B [80]. One way to achieve stability is to introduce some mechanism that couples the rate of depolymerization V_{depot} to the sliding force F_{slide}. For example, Goshima et al. [233] propose that increases in F_{slide} push the minus ends of MTs closer to the centrosome where depolymerizing motors act. Using the theory of Brownian ratchets (Sect. 4.2), the probability that an MT is within a critical distance δ of the centrosome is $P(x < \delta) = 1 - \exp(-F_{\text{slide}}\delta/Nk_BT)$. It is assumed that the sliding force is divided equally by the N MTs attached to the centrosome. Assuming that depolymerization by motors only occurs if $x < \delta$, then

$$V_{\text{depol}} = V_0 + V_1\left(1 - \mathrm{e}^{-F_{\text{slide}}\delta/Nk_BT}\right). \tag{8.2.32}$$

Combining these various equations we obtain the pair of equations

$$\frac{dL}{dt} = 2(V_{\text{poly}} - V_{\text{slide}}) \tag{8.2.33a}$$

$$\mu\frac{dS}{dt} = 2\left(\alpha L\left(1 - \frac{V_{\text{slide}}}{V_{\max}}\right) - F_{\text{Kt}} - \beta(S - S_0)\right), \tag{8.2.33b}$$

with

$$V_{\text{slide}} = \frac{1}{2}\frac{dS}{dt} + V_0 + V_1\left(1 - \mathrm{e}^{-\alpha L\left(1-\frac{V_{\text{slide}}}{V_{\max}}\right)\delta/NK_bT}\right).$$

The steady-state solution for S and L are then (see Ex. 8.5)

$$L = \frac{\Gamma_0 k_B TN}{\alpha\delta(V_{\text{poly}}/V_{\max} - 1)}, \quad S = S_0 - \frac{F_{\text{Kt}}}{\beta} - \frac{\Gamma_0 Nk_BTN}{\beta\delta}, \tag{8.2.34}$$

where

$$\Gamma_0 = \ln\left(1 - \frac{V_{\text{poly}} - V_0}{V_1}\right).$$

8.3 Cell Motility

Just as the polymerization and depolymerization of microtubules plays an essential role in cell mitosis, the growth and shrinkage of actin polymers plays a major role in generating the forces necessary for various forms of cell motility [138, 446, 512, 520]. For example, the movement of crawling cells such as amoeba, keratocytes, fibroblasts, and migrating neurons involves the protrusion of lamellipodia and filopodia at the leading edge of the cell (see Fig. 8.18a), which requires actin polymerization at the cell membrane boundary [61]. On the other hand, intracellular pathogens such as *Listeria* propel themselves within a host cell by assembling the host cell's actin into a comet-like tail. The tail consists of oriented cross-linked networks of actin filaments whose growing ends orient towards the bacterial surface,

thus thrusting the pathogen forward [626, 627] (see Fig. 8.18b). A major challenge is linking the complex biochemical processes regulating actin polymerization with mechanical properties of the cell and the associated forces. In the case of crawling cells, there are contractile forces on the actin cytoskeleton due to the action of myosin motors, traction forces from the drag between the cytoskeleton and surface adhesion complexes, membrane tension resisting the actin polymerization force, viscoelastic stresses arising from deformations of the actin network, and viscous drag between actin filaments and cytosolic fluid flows [134, 371, 444]. As in the case of cell mitosis, cell motility is a vast subject in its own right and we cannot hope to do it justice here. Instead, we will focus on some aspects that relate most closely to the themes of this book. For an excellent introduction to cell motility see the book *Cell movements: from molecules to motility* by Bray [61].

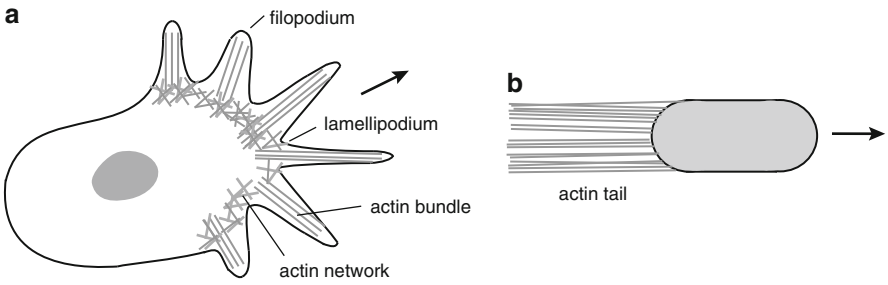


Fig. 8.18: Examples of actin-based cell motility. (**a**) Crawling eukaryotic cell. (**b**) Pathogen such as *Listeria* propelled by an actin tail assembled from the cytoskeleton of the host cell

8.3.1 Tethered Ratchet Model

We begin by considering a microscopic model of cell protrusion based on an extension of the polymerization ratchet model [507]. This was developed by Mogilner and Oster within the context of the simpler problem of *Listeria* propulsion [447, 448] and has subsequently been incorporated into more complex models of cell crawling [270, 329, 446]. Recall from Sect. 4.2 that the speed of growth of a polymerization ratchet depends on the diffusion coefficient of membrane Brownian motion. Within the context of bacterial motion, this would imply that the bacterial velocity depends on its diffusion coefficient and thus on its size. However, such size dependence has not been observed experimentally, which led Mogilner and Oster to propose an *elastic ratchet* model, in which thermal bending fluctuations of a semi-stiff actin filament, rather then bacterial diffusion [447], generates the gap necessary for insertion of an additional monomer, with the resulting growth generating the force to propel the bacterium forward. The elastic ratchet model was itself superseded by

the *tethered ratchet* model [448], in order to account for a number of additional experimental observations [91]. In particular, Brownian fluctuations are almost completely suppressed during *Listeria* propulsion due to the fact that the bacterium is tightly bound to its actin tail. One thus observes smooth particle trajectories that are persistent in both direction and curvature. It is known that the surface of *Listeria* is coated with nucleation promotion factor ActA, which transiently binds the Arp2/3 complex on the sides of attached actin filaments; Arp2/3 is known to mediate nucleation of side-branched filaments. The tethered ratchet model is one way to resolve the dilemma of how the actin tail can be attached to the bacterium and yet propel the bacterium forward via growth of unattached active filaments. More specifically, it proposes that there are two classes of filament: some are attached, under tension and nucleating rather than growing, while others are unattached and pushing via an elastic ratchet mechanism. In the following, we will describe the tethered ratchet model in more detail.

Suppose that there are $n_u(t)$ unattached filaments and $n_a(t)$ attached filaments at time t and that the bacterium is moving at speed v. There are three forces acting on the bacterium, neglecting any elastic recoil forces of the actin tail (see Fig 8.19): a load force $F_L = \gamma v + F_{\text{ext}}$, where γ is a viscous drag coefficient and F_{ext} represents any experimentally imposed external forces; a tensional force $F_a = n_a f_a$ due to attached filaments, with f_a the force per filament; a pushing force $F_u = n_u f_u$ due to unattached filaments, with f_u the force exerted by a single unattached filament via an elastic ratchet mechanism. The corresponding force-balance equation is

$$F_L + n_a f_a = n_u f_u. \tag{8.3.1}$$

It is assumed that the two filament populations evolve according to the simple kinetic equations

$$\frac{dn_a}{dt} = \sigma - k n_a, \quad \frac{dn_u}{dt} = k n_a - \kappa n_u, \tag{8.3.2}$$

where σ is the nucleation rate of side branches, k is the rate of detachment, and κ is the rate of capping of unattached filaments which can then no longer polymerize and push on the bacterium's surface. It remains to specify the dependence of the forces f_a, f_u and the detachment rate k on the bacterium velocity v. The force-velocity relation for a single polymerizing filament is taken from the polymerization ratchet model [447] (see equation (4.2.12)):

$$v = v_+ e^{-f_u l / k_B T} - v_-, \tag{8.3.3}$$

where $v_+ = k_{\text{on}} l M$ is the polymerization velocity and $v_- = k_{\text{off}} l$ is the depolymerization velocity. Here k_{on} and k_{off} are the rates of monomer assembly and disassembly, M is the concentration of monomers available for polymerization, and l is the effective increase in filament length due to addition of one monomer.

In order to estimate the average attachment force f_a, it is assumed that an attached filament acts like a Hookean spring. Suppose that the filament binds to an ActA complex at time $t = 0$. The force acting on the resulting bond is given by $f(t) = \eta v t$,

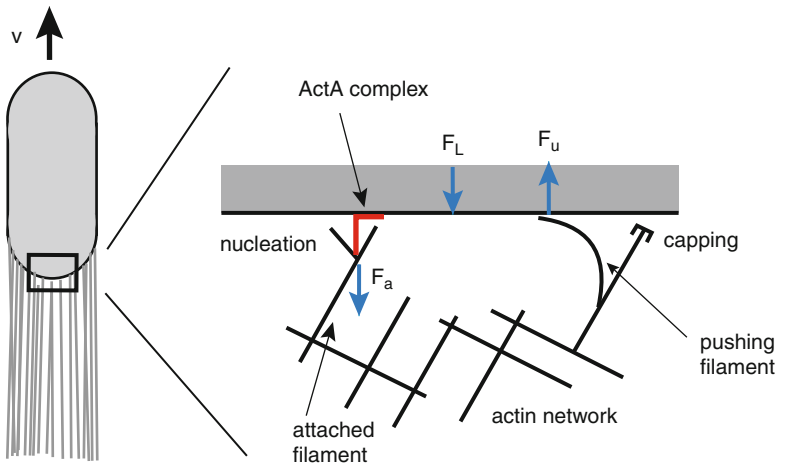


Fig. 8.19: Schematic diagram of tethered ratchet model, redrawn from [448]. There are three main forces acting on the Listeria: a load force F_L due to viscous drag and any externally imposed forces (in an experiment), a polymerization ratchet force F_u due to unattached filaments, and a tensional force F_a due to attached filaments with $F_L + F_a = F_u$

where η is the effective spring constant. Using the basic theory of chemical bond breaking [175], the rate of detachment takes the velocity-dependent form

$$k(v,t) = k_0 \mathrm{e}^{\eta v t / f_b}, \tag{8.3.4}$$

where $x_b = f_b/\eta$ can be interpreted as bond length at which the bond breaks sharply. The probability $p(t)dt$ of the bond first breaking in the time interval $(t, t+dt)$ is given by the product of no failure in the interval $(0,t)$ times the probability of subsequent failure within the interval $(t, t+dt)$. Hence,

$$p(t) = k(v,t)\mathrm{e}^{-\int_0^t k(v,s)ds}.$$

Set $v_0 = f_b k_0/\eta$, which can be interpreted as the velocity at which the bond stretches to the critical length x_b over the characteristic bond lifetime $1/k_0$. Rescaling velocity and time according to $\mu = v/v_0$ and $\tau = k_0 t$, we have

$$p(\tau) = \exp\left(\mu\tau + \frac{1}{\mu}(1 - \mathrm{e}^{-\mu\tau})\right).$$

It follows that the mean attachment time of a filament (for constant v) is

$$\langle t \rangle = \frac{1}{k_0}\int_0^\infty \tau p(\tau) d\tau = \frac{1}{k_0} w(\mu), \tag{8.3.5}$$

with

$$w(\mu) = \int_0^\infty \tau \exp\left(\mu\tau + \frac{1-\mathrm{e}^{-\mu\tau}}{\mu}\right) d\tau.$$

We now identify the mean detachment rate as $k = 1/\langle t\rangle$ and take the average force f_a exerted by a single attached filament to be $f_a = \eta v\langle t\rangle$. Thus,

$$k(\mu) = \frac{k_0}{w(\mu)}, \quad f_a = f_b \mu w(\mu). \tag{8.3.6}$$

Note that the function $w(\mu)$ has the following properties:

1. If $\mu \ll 1$, then $w(\mu) \approx 1$, which implies that for sufficiently slow movement ($v \ll v_0$) the effective detachment rate is equal to the force-free rate ($k \approx k_0$) and $f_a \approx f_b v/v_0$.
2. If $\mu \gg 1$, then $w(\mu) \approx \mu^{-1} \ln \mu$.

Consider the case of constant propulsion speed v. The steady-state numbers of attached and detached filaments are then

$$n_a = \sigma/k, \quad n_u = \sigma/\kappa.$$

Substituting the force-balance equation (8.3.1) into the velocity equation (8.3.3) gives

$$v = v_+ \exp\left[-l(n_a f_a/n_u + F_l/n_u)/k_B T\right] - v_-.$$

Since $n_a/n_u = \kappa/k = w(\mu)\kappa/k_0$ and $f_a = f_b \mu w(\mu)$, the velocity satisfies the implicit equation

$$v = v_+ \exp\left[-l\left(\frac{f_b k_0}{\kappa v_0} v w^2(v/v_0) + \frac{F_L \kappa}{\sigma}\right)/k_B T\right] - v_-. \tag{8.3.7}$$

Using biophysically based estimates for the various parameters, Mogilner and Oster [448] numerically solved the equation for v and obtained speeds of the order 10 nm /s, which is consistent with experimental data (see Fig. 8.20). They also showed that the load–velocity relation exhibits biphasic behavior, whereby the velocity decreases rapidly with F_L at low load forces and decreases more slowly at high load forces. This is a consequence of the modeling assumptions regarding chemical bond breaking. At high velocities increasing the external load helps the attached filaments to hold on longer, thus increasing the resistive force F_a which itself slows the bacterium further. On the other hand, at sufficiently slow velocities the external load has a minor effect on the resistive force F_a and the velocity decreases more slowly.

The tethered ratchet model has been refined over the years in order to include more details regarding nucleation and capping mechanisms, for example, and to try and account for an ever-increasing amount of new biophysical data [444]. Several other propulsion mechanisms have also been proposed, which can match some types of data better than ratchet models. One alternative model assumes that all filaments are attached to the surface of the bacterium, with the pushing (barbed)

ends clamped to an end-tracking protein at the surface, which processively tracks the growing filament tip via phosphorylation of ATP [138, 139]. Another class of

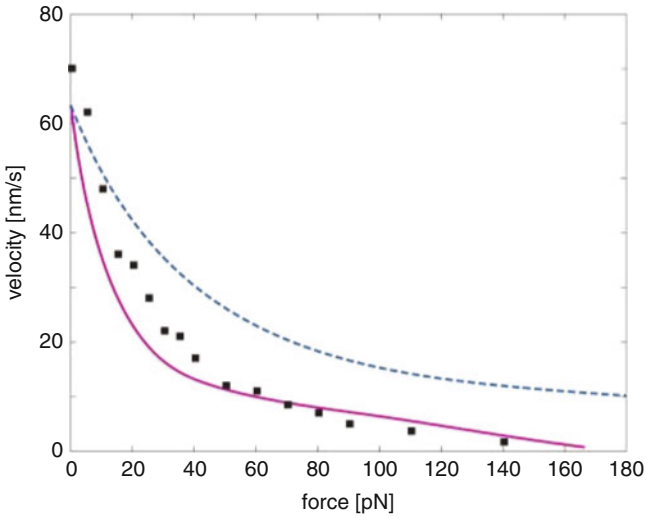


Fig. 8.20: The load–velocity curve for the tethered ratchet model of Mogilner and Oster [448]. The *solid curve* is generated from equation (8.3.7) using the following parameter values: monomer size $l = 2.2$ nm, polymerization velocity $v_+ = 500\ \text{nms}^{-1}$, depolymerization velocity $v_- = 2.2\ \text{nms}^{-1}$, nucleation rate $\sigma = 10\ \text{s}^{-1}$, capping rate $\kappa = 0.5\ \text{s}^{-1}$, free detachment rate $k_0 = 0.5\ \text{s}^{-1}$, thermal energy $k_B T = 4.1\text{pN} \cdot \text{nm}$, effective length of bond $x_b = 0.4$ nm, effective strength of bond $f_b = 10$ pN, and spring coefficient $\eta = 1\text{pN/nm}$. The *dashed curve* is obtained by introducing a threefold increase in the nucleation rate σ and illustrates the effect of filament density on the load–velocity behavior. Finally, the *squares* represent data from stochastic model simulations with a reduced polymerization velocity $v_+ = 240\ \text{nms}^{-1}$ (Adapted from Mogilner and Oster [448])

model is based on the analysis of elastic deformations of an actin gel near the surface of the bacterium [211, 416]. The gel is approximated as a continuous elastic medium with stress generated at the surface interface by growing actin filaments. One finds that near curved surfaces, squeezing elastic forces propel the bacterium forward.

8.3.2 Crawling Cells and Lamellipodial Motility

The majority of migratory cells that crawl along some cellular substrate such as the extracellular matrix rely on lamellipodial motility. A prerequisite for directed motion is that the cell is polarized, i.e., it has a well-defined front and rear. (Mechanisms of cell polarization for a range of cellular processes will be considered in Sect. 9.5.) The leading edge of a crawling cell consists of a lamellipod, which is

a flat leaf-like extension that is filled with a dense actin network. Roughly speaking, following cell polarization, the migration of a cell consists of three components driven by interconnected but distinct processes [134, 444] (see also Fig. 8.21):

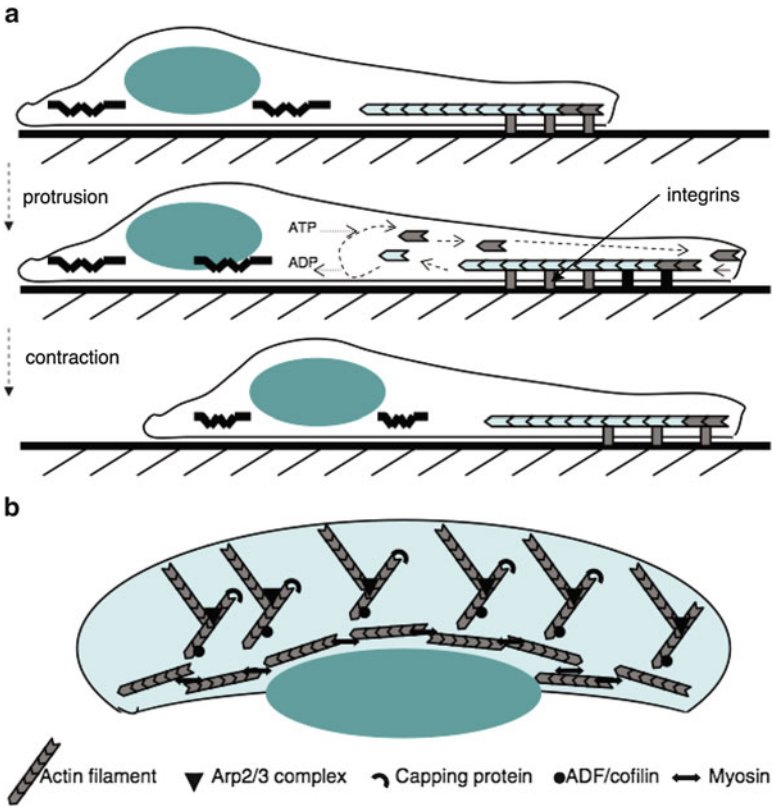


Fig. 8.21: Schematic illustration of main components of cell crawling. (**a**) *Side* view shows protrusion at the leading edge due to polymerization of the actin network followed by contraction or retraction of the rear due to action of myosin II motors. Net displacement occurs proving that there is a gradient of cell adhesion from front to rear. (**b**) *Top* view of a crawling cell showing the actin-rich lamellipod whose branching actin network is modifies by various actin regulatory proteins such as Arp2/3 (nucleation promoter and ADF/cofilin (capping promoter) Adapted from Mogilner [444]

1. *Cell protrusion:* Models of cell protrusion combine ratchet models of actin polymerization with the so-called dendritic-nucleation hypothesis. The latter posits that nascent actin filaments branch from the sides of existing filaments such that there is approximately a 70° angle between "mother" and "daughter" filaments, and all leading edge filaments have their barbed ends oriented towards the direction of protrusion at an angle of around 35° [96, 410, 458, 568]. One finds that the growing barbed ends grow at a rate around 0.1 μm/sec, thus pushing out the

membrane of the leading edge. However, capping of the growing ends restricts the length of individual filaments to $0.1 - 1\ \mu$m. These features make sense, since intermediate-length filaments that subtend the leading edge at an angle are neither too rigid nor too flexible and are protected from immediate capping. The net protrusion of the leading edge is determined by actin polymerization minus the centripetal rearward flow of the actin network. This retrograde flow is a dissipative mechanism for reducing stresses on the actin network.

2. *Cell contraction:* Contraction of the rear of the cell is thought to be mediated by myosin II motors acting on actin fibers in an analogous fashion to muscle contraction. A gradient of adhesion forces that is high at the leading edge and low at the trailing edge (see below) means that contraction will lead to preferential forward movement of the rear, provided that it is not too strong that it cancels protrusion by imposing a rearward stress on the actin network.

3. *Cell adhesion:* Recall from our discussion of the mitotic spindle (Sect. 8.2) that inertial effects can be ignored in the low Reynolds number regime, so there has to be force balance. In order to balance propulsive forces at the leading edge and contractile forces at the rear, adhesion forces between the cell and extracellular matrix are required [500]. The latter forces are mediated by transmembrane receptors known as *integrins* [656]. At the leading edge, the adhesion acts like a clutch [101]. That is, for high adhesion or drag, the retrograde flow is slow and polymerization results in a net protrusion—the clutch is in "drive." On the other hand, if adhesion is weak, then retrograde flow can cancel the polymerization, and the actin network treadmills, i.e., the clutch is in "neutral." There is growing experimental evidence that in addition to acting as mechanical linkages, integrins couple to various signaling pathways that modulate actin polymerization and activate myosin motors [54, 203]. Moreover, protrusion appears to be coupled to the formation of nascent integrins at the leading edge [110], which subsequently disassemble towards the rear resulting in an adhesion gradient. The existence of an adhesion gradient was assumed in one of the earliest whole-cell mechanical models of cell motility [141].

8.3.3 Stochastic Model of Integrin-Mediated Cell Protrusion

There are a wide variety of mathematical and computational models of cell migration; see the recent reviews by Mogilner and collaborators [134, 444]. Most models focus on the three main processes of cell migration—protrusion, contraction, and adhesion. We will consider one recent example, namely, a stochastic model of cell protrusion at the leading edge, which takes into account the formation of nascent integrins and their coupling to the actin cytoskeleton and myosin motors [676] (see Fig. 8.22). Following Welf et al. [676], we present the deterministic version of the model and then indicate how fluctuations are incorporated. For simplicity, the leading edge of the cell is treated as a single homogeneous compartment

with $m(t)$ the density of active myosin motors, $a(t)$ the density of nascent adhesion molecules, and $r(t)$ the density of Rac GTPases, which act as signaling molecules. The velocity V_p of polymerization is assumed to depend on the concentration of Rac

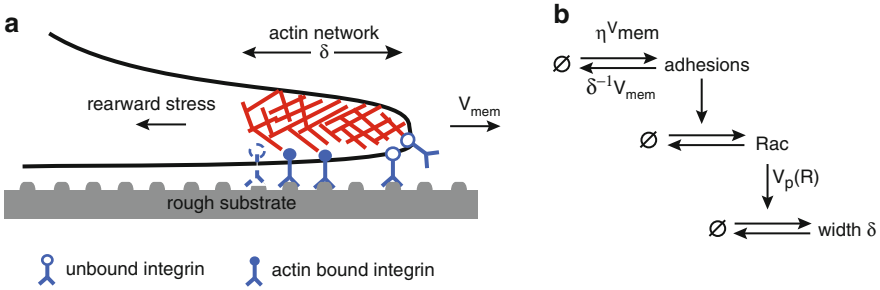


Fig. 8.22: Model of integrin-based leading edge protusion [676]. (**a**) Cartoon of nascent adhesion formation and binding. Adhesions are formed at a rate ηV_{men}, where V_{men} is the protrusion velocity, and bind to the actin network. Turnover of adhesions towards the back occurs at a rate $k_a + V_{\text{men}}/\delta$, where δ is the width of the actin network. (**b**) Schematic of reaction network for coupling between adhesion and actin mediated by Rac

according to a sigmoid-like function, whereas the rate of depolymerization is taken to be proportional to the width δ of the lamellipodium. Thus

$$\frac{d\delta(t)}{dt} = V_{\text{p}}(r(t)) - k_d\delta(t), \quad V_{\text{p}}(r) = \frac{V_{\text{p,o}} + V_{\text{p,max}}K_v r}{1 + K_v r}, \tag{8.3.8}$$

where k_d is a rate constant. Activation of Rac is mediated by the nascent adhesions according to the simple first-order kinetic scheme

$$\frac{dr(t)}{dt} = k_+ a(t) - k_- r(t), \tag{8.3.9}$$

whereas the rates of production and degradation of adhesions depends on the velocity V_{mem} of membrane protrusion:

$$\frac{da(t)}{dt} = \eta V_{\text{mem}}(t) - (k_a + V_{\text{mem}}(t)/\delta)a(t), \tag{8.3.10}$$

Here k_a is a background degradation rate of adhesions, and η is an adjustable parameter that determines the efficiency of adhesion formation at the leading edge as a function of the extracellular matrix density. The relationship between protrusion velocity and the speed of polymerization is taken to be [23]

$$V_{\text{mem}} = V_{\text{p}} - V_{\text{ret}}, \tag{8.3.11}$$

where $V_{\rm ret}$ is the retrograde flow velocity of the actin network. For fixed $V_{\rm ret}$, equations (8.3.8)–(8.3.10) represent a protrusion-based positive feedback loop: an increase in nascent adhesion strengthens the Rac signal, which results in faster polymerization and growth of $\delta(t)$, and this in turn increases the production rate of nascent adhesions.

In order to determine $V_{\rm ret}$, it is necessary to consider a force-balance equation, which is assumed to be of the form [676]

$$\alpha V_{\rm mem}(t) + \beta m(t) = \sigma_c(t) + \xi V_{\rm ret}(t), \tag{8.3.12}$$

where $\alpha V_{\rm men}$ is the rearward stress due to resistance of the membrane to protrusion, which is assumed to be proportional to the protrusion velocity, $\beta m(t)$ is the rearward stress due to the action of myosin motors, σ_c is the opposing forward stress exerted by the adhesion-based clutch, and $\xi V_{\rm ret}$ is the dissipation of stress due to retrograde flow. Here α, β, and ξ are constants. The next assumption is that the distribution of stress between the clutch and retrograde flow depends on the density a_c of adhesion molecules bound to the actin network,

$$\frac{\sigma_c}{\sigma_c + \xi V_{\rm ret}} = \frac{\alpha_c a_c}{1 + \alpha_c a_c}. \tag{8.3.13}$$

Assuming that bound and unbound adhesions are in thermodynamic equilibrium, the fraction of bound adhesions is given by a Boltzmann–Gibbs distribution,

$$\frac{a_c}{a} = \frac{K_c \mathrm{e}^{-f/f_c}}{1 + K_c \mathrm{e}^{-f/f_c}}, \tag{8.3.14}$$

where $f = \sigma_c / N_c$ is the force on a single adhesive bond. (A similar exponential dependence of unbinding rates on the applied force was assumed in the ToW model of collective motor transport; see Sect. 4.4.2.) The final assumption of the model is that the adhesive clutch activates myosin according to the first-order kinetics

$$\frac{dm(t)}{dt} = \gamma_+ a_c^*(t) - \gamma_- m(t), \tag{8.3.15}$$

where $a_c^*(t)$ is the density of bound adhesions that activate myosin. The latter is also assumed to be determined by a Boltzmann–Gibbs distribution,

$$\frac{a_c^*}{a_c} = \frac{K_c^* \mathrm{e}^{-f/f_c^*}}{1 + K_c^* \mathrm{e}^{-f/f_c^*}}. \tag{8.3.16}$$

Welf et al. [676] performed computer simulations on a stochastic version of the above system of equations, which takes into account intrinsic fluctuations in the number of each molecular species, namely, integrins, active myosin motors, and Rac GTPases. The construction of the master equation follows along identical lines to the various gene networks considered in Chap. 6. The resulting master equation is then simulated using a Gillespie-based algorithm. Details of the simulations and

parameter values can be found in [676]. One of the main findings of the stochastic model is that the protrusion velocity exhibits biphasic behavior, tending to switch rapidly between on and off states, and that the distributions of times spent in the protrusion phase have a characteristic distribution that is consistent with experimental observations. Moreover, the distribution of protrusion durations shifts significantly to shorter times as the density of the extracellular matrix is increased (larger η)

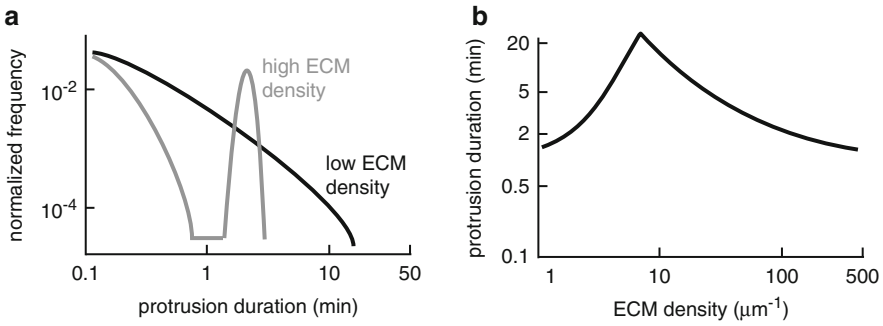


Fig. 8.23: Sketch of results from computer simulations of the stochastic version of the integrin-based model carried out by Welf et al. [676]. (**a**) Distributions of protrusion durations for low and high extracellular matrix densities. (**b**) Mean protrusion duration as a function of ECM density, showing a peak where the ECM density is optimal for cell motility

(see Fig. 8.23). In the model, this is a consequence of a higher rate of activation of adhesions, which then activate more myosin's that oppose protrusion. Note that the peak in the distribution of duration times at high ECM density reflects a strong deterministic oscillatory component of the profusion dynamics.

Finally, note that there have also been a number of integrative models of the whole cell, which can make predictions about the changes in cell shape during migration [270, 329, 417]. Mathematically speaking, the modeling of cell shape involves a difficult free boundary value problem, in which the cell is represented as a time-dependent domain $\Omega(t)$ whose boundary $\partial\Omega(t)$ evolves according to boundary conditions that themselves depend on various densities that are defined by dynamical equations on $\Omega(t)$. We refer the reader to the given citations for more details.

8.4 Cooperative Transport of Proteins Between Cellular Organelles

The extensive secretory pathway of eukaryotic cells is a critical system for the maturation and transportation of newly synthesized lipids and proteins to specific target sites within the cell membrane. The first major organelle of the secretory pathway is

the endoplasmic reticulum (ER) [393] (see Fig. 8.24). Proteins and lipids destined for the plasma membrane enter the ER from the nucleus as they are translated by ER-associated ribosomes, where they fold into their proper 3D structure. One important aspect of the secretory pathway is that it is tightly regulated [393]. Proteins accumulate at specific exit sites and leave the ER in vesicles that transfer the cargo to organelles forming the Golgi network, where final packaging and sorting for target delivery is carried out. In most eukaryotic cells the Golgi network is confined to a region around the nucleus known as the Golgi apparatus. Another of the significant features of the secretory pathway is that there is a constant active exchange

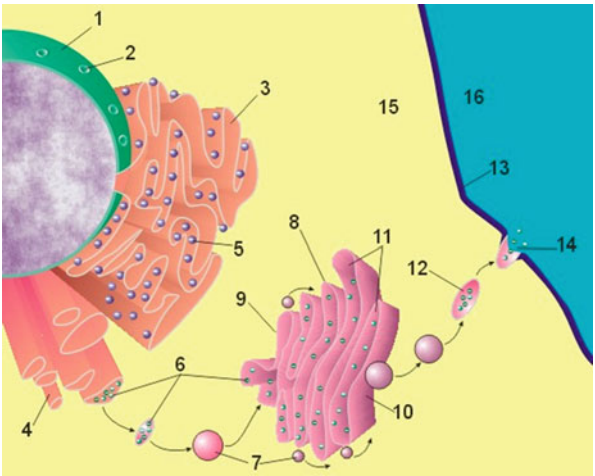


Fig. 8.24: Diagram of secretory pathway including nucleus, ER, and Golgi apparatus. 1. Nuclear membrane. 2. Nuclear pore. 3. RER. 4. SER. 5. Ribosome. 6. Protein. 7. Transport vesicles. 8. Golgi apparatus. 9. *Cis* face of Golgi apparatus. 10. *Trans* face of Golgi apparatus. 11 Cisternae of Golgi apparatus. 12. Secretory vesicle. 13. Plasma membrane. 14. Exocytosis. 15 Cytoplasm. 16. Extracellular domain (Public domain image from WikiMedia Commons)

of molecules between organelles such as the ER and Golgi apparatus, which have different lipid and protein compositions. Such an exchange is mediated by motor-driven vesicular transport. Vesicles bud from one compartment or organelle, carrying various lipids and proteins, and subsequently fuse with another compartment. Transport in the anterograde direction from the ER to Golgi has to be counterbalanced by retrograde transport in order to maintain the size of the compartments and to reuse components of the transport machinery. Since bidirectional transport would be expected to equalize the composition of both compartments, there has been considerable interest in understanding the self-organizing mechanisms that allow such organelles to maintain their distinct identities while constantly exchanging material [441]. One transport model for generating stable, nonidentical compartments has been proposed by Heinrich and Rapoport [256], and developed further by a number

of groups [51, 144, 226]. These models primarily focus on two of the essential steps of vesicular transport, namely budding and fusion, both of which involve a complex network of molecular interactions between vesicles, transported molecules, and recipient organelles [337, 392]. Such nonlinear cooperative interactions provide the basic mechanism for the self-organization of distinct organelles, and, hence, we refer to the whole process as an example of cooperative vesicular transport.

Budding from a donor compartment is mediated by cytosolic protein coats that bind to the membrane, induce curvature, and eventually pinch off a vesicle. Protein coats, which are recruited to the membrane by the hydrolysis of G-proteins, are also involved in the selective concentration of specific proteins within the budding vesicle. Following budding and detachment from the donor compartment, vesicles are transported to the acceptor compartment where they undergo fusion. The latter is mediated by the pair-wise interaction of vesicle ($v-$) and target ($t-$) *soluble N-ethylmaleimide-sensitive factor attachment protein receptors* (SNARES). Although $v-$snares and $t-$snares occur in multiple homologous variants, they tend to form high-affinity pairs that provide a recognition mechanism for membrane fusion between specific vesicles and a compartment. Since each type of protein coat preferentially loads particular high-affinity pairs of SNARES during budding, it follows that a given vesicle fuses preferentially with a compartment that contains higher levels of these SNARES, thus further increasing their concentration within the compartment. In the case of a single type of protein coat mediating vesicular transport between two compartments, this mechanism would lead to a steady-state in which many vesicles with a low SNARE content move in one direction (say, from the first to the second compartment), whereas a few vesicles with a large SNARE content move in the opposite direction. The total protein fluxes would then be balanced, while maintaining a higher concentration of SNARES in the second compartment. However, lipid balance would not be maintained because there would be a net flux of vesicles in the anterograde direction. The simultaneous balance of lipid fluxes could be achieved by having a second type of protein coat that preferentially loads a different set of SNARES that are concentrated in the first compartment, resulting in a net flux of vesicles in the retrograde direction. The role of protein coats and SNARES in vesicular budding and fusion forms the basis of the model introduced by Heinrich and Rapoport [256]. Note that in the case of the early secretory pathway coat protein complex II (COPII) vesicles mediate transport from the ER to the Golgi, whereas coat protein complex I (COPI) vesicles mediate transport in the opposite direction.

We will describe a simplified version of the model by Rapoport and Heinrich [256] see [65]. Suppose that there are two compartments $j = 1,2$ that contain two types of surface protein $X = U,V$ whose concentrations at time t are denoted by $U_j(t), V_j(t)$, respectively (see Fig. 8.25). For simplicity, we do not explicitly model the interactions between complementary SNARE proteins, that is, we take each protein U,V to be self-interacting. These proteins are exchanged between the compartments via two types of protein-coated vesicles labeled by $\alpha = a,b$ with a-vesicles transporting protein U and b-vesicles transporting protein V. As a further simplification, we neglect the actual physical process of vesicular transport and the associated

transport delays. Thus, we assume vesicles that bud from the jth compartment are immediately available for fusion with either compartment. Let $Q_a(U_j)$ denote the backward rate per unit area of fusion of an a-vesicle with its source compartment, whereas $P_a(U_j, U_{\bar{j}})$ is the forward rate per unit area at which an a-vesicle from compartment $j = 1,2$ fuses with compartment $\bar{j} = 2,1$. The backward and forward rates for a-vesicles are taken to have the explicit form [256]

$$Q_a(U_j) = \gamma + \kappa f_a(U_j) U_j, \tag{8.4.1}$$

$$P_a(U_j, U_{\bar{j}}) = \gamma + \kappa f_a(U_j) U_{\bar{j}}. \tag{8.4.2}$$

Here $f_a(U)$ is the protein and coat specific concentration within an a-vesicle generated from a compartment with concentration U. It is taken to have the Michaelis-Menten form

$$f_a(U) = W_a \frac{U}{U + C_a}. \tag{8.4.3}$$

(In a more general model, vesicles would transport both proteins in a competitive manner [71, 144, 226, 256].) Note that γ can be interpreted as a background fusion rate, whereas κ determines the increase in the reaction rate due to binary interactions between surface proteins of the vesicle and compartment, respectively, based on the law of mass action. For simplicity, κ is taken to be independent of the particular protein. Analogous equations hold for b-vesicles with $a \to b$ and $U \to V$.

Let $N_j^\alpha(t)$ denote the number density of α-vesicles produced by the jth compartment at time t, $\alpha = a, b$. The number of vesicles evolves according to the equation

$$\frac{dN_j^a}{dt} = A_j K_a - A_j Q_a(U_j) N_j^a - A_{\bar{j}} P_\alpha(U_j, U_{\bar{j}}) N_j^a. \tag{8.4.4}$$

Here $A_j(t)$ is the surface area of the jth compartment and K_α is the rate of production of α-vesicles. Assuming that all vesicles have the same surface area ΔA with $\Delta A \ll A_j$, the rate of change of the number of U proteins within the jth compartment is

$$\frac{d[A_j U_j]}{dt} = A_j \Delta A \left[-K_a + Q_a(U_j) N_j^a + P_a(U_{\bar{j}}, U_j) N_{\bar{j}}^a \right] f_a(U_j). \tag{8.4.5}$$

We also have a conservation equation for the total number M_U of U proteins:

$$M_U = \sum_{j=1,2} \left(A_j(t) U_j(t) + \Delta A \, f_a(U_j) N_j^a(t) \right). \tag{8.4.6}$$

Identical equations hold for b-vesicles and V proteins by taking $a \to b$ and $U \to V$ in equations (8.4.4), (8.4.5), and (8.4.6). In this simplified model, competition between the proteins arises via the transport of membrane. The rate of change of membrane surface area of the jth compartment is

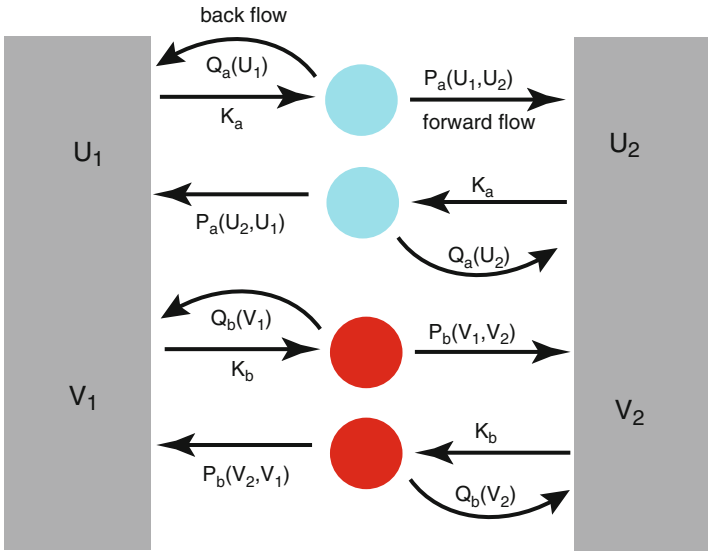


Fig. 8.25: Schematic diagram of the exchange of vesicles between two organelles. Shown is the limiting case in which one class of vesicle (*shaded blue*, labeled $\alpha = a$) only carries protein U and the other type of vesicle (*shaded red*, labeled $\alpha = b$) only carries protein V. When a vesicle of type α buds from a given compartment, it can immediately re-fuse with the same compartment at a rate Q_α or fuse with the other compartment at a rate P_α. Both compartments produce vesicles of type α at a rate K_α. The nonlinear interactions between vesicles and compartments can generate a symmetry-breaking bifurcation, in which the two compartments are maintained at different protein concentrations

$$\frac{d[A_j]}{dt} = A_j \Delta A \left[-K_a + Q_a(U_j) N_j^a + P_a(U_{\bar{j}}, U_j) N_{\bar{j}}^a \right] \tag{8.4.7}$$

$$+ A_j \Delta A \left[-K_b + Q_b(V_j) N_j^b + P_b(V_{\bar{j}}, V_j) N_{\bar{j}}^b \right], \tag{8.4.8}$$

which is supplemented by the conservation condition for the total amount of membrane A_{tot}

$$A_{\text{tot}} = \sum_{j=1,2} \left(A_j(t) + \Delta A [N_j^a(t) + N_j^b(t)] \right). \tag{8.4.9}$$

We will focus on steady-state solutions. Setting to zero all time derivatives in equation (8.4.4) with U_j, V_j fixed, immediately implies that at steady state

$$N_j^a = \frac{A_j K_a}{A_j Q_a(U_j) + A_{\bar{j}} P_a(U_j, U_{\bar{j}})}, \quad N_j^b = \frac{A_j K_b}{A_j Q_b(V_j) + A_{\bar{j}} P_b(V_j, V_{\bar{j}})}. \tag{8.4.10}$$

Setting $d[A_j X_j]/dt = 0$ in equation (8.4.5) and substituting the steady-state expressions for N_j^a, N_j^b gives

$$G_a(U_1,U_2)f_a(U_1) = \hat{G}_a(U_1,U_2)f_a(U_2) \tag{8.4.11}$$

$$G_b(V_1,V_2)f_b(V_1) = \hat{G}_b(V_1,V_2)f_b(V_2) \tag{8.4.12}$$

where

$$G_a(U_1,U_2) = \frac{P_a(U_1,U_2)}{A_1Q_a(U_1)+A_2P_a(U_1,U_2)}, \hat{G}_a(U_1,U_2) = \frac{P_a(U_2,U_1)}{A_2Q_a(U_2)+A_1P_a(U_2,U_1)}, \tag{8.4.13}$$

and similarly for $a \to b$, $U \to V$. The steady-state version of (8.4.7) yields

$$K_aG_a(U_1,U_2)+K_bG_b(V_1,V_2) = K_a\hat{G}_a(U_1,U_2)+K_b\hat{G}_b(V_1,V_2). \tag{8.4.14}$$

It turns out that the system undergoes a symmetry-breaking bifurcation even if the kinetics of both proteins are the same, that is, $W_a = W_b = W$, $C_a = C_b$, and $K_a = K_b = K$. In this case the functions $f_\alpha, P_\alpha, Q_\alpha$, $G_\alpha, \hat{G}_\alpha$ become independent of the

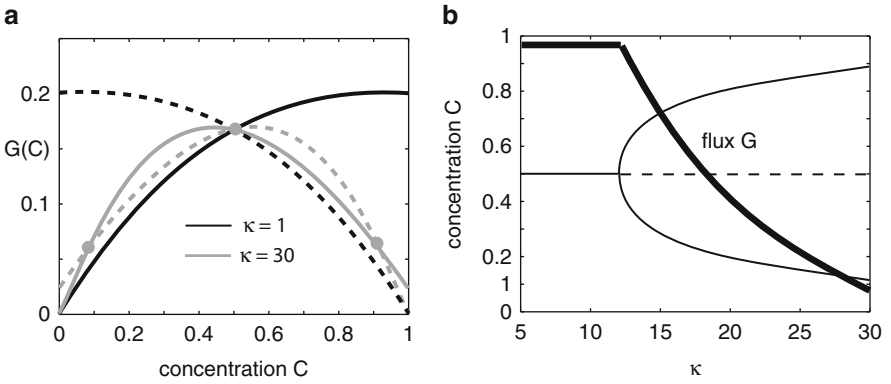


Fig. 8.26: Symmetry breaking in two-compartment model. (**a**) Plot of functions $G(C)$ (*solid curves*) and $G(C_{\rm tot}-C)$ (*dashed curves*) for various values of κ. Points of intersection of $G(C)$ and $G(C_{\rm tot})$ for fixed κ determine steady states. Other parameter values are $\gamma = 1$, $W = 1$, $C_{\rm S} = 1$, $A_{\rm tot} = 1$, and $C_{\rm tot} = 1$. (**b**) Bifurcation diagram for steady-state concentration C in compartment 1 as a function of the parameter κ. *Thick curve* shows the variation of the steady-state flux $G(C)$ along a stable branch

label $\alpha = a,b$. In this simplified case, one solution to equations (8.4.11)–(8.4.14) is the symmetric solution

$$(U_1,V_1,A_1) = (U_2,V_2,A_2) = (U^*,V^*,A^*),$$

with (U^*,V^*,A^*) determined from the conservation conditions (8.4.9) and (8.4.6). In order to search for symmetry-breaking solutions (nonidentical protein concentra-

tions in the two compartments), suppose that $A_1 = A_2 = A_{\text{tot}}/2$ (equal surface areas) and

$$M_U = M_V \equiv M. \tag{8.4.15}$$

Assuming that $\varepsilon = \Delta A/A_{\text{tot}} \ll 1$, the conservation conditions (8.4.9) and (8.4.6) imply that

$$A_2 \approx A_{\text{tot}} - A_1, \quad A_2 X_2 \approx M - A_1 X_1 \tag{8.4.16}$$

for $X = U,V$, and the approximate solution of equations (8.4.11), (8.4.12), and (8.4.14) is

$$U_1 = C, \quad U_2 = C_{\text{tot}} - C, \quad V_1 = C_{\text{tot}} - C, \quad V_2 = C, \tag{8.4.17}$$

with $C_{\text{tot}} = 2M/A_{\text{tot}}$ and C satisfying the equation

$$\frac{P(C,C_{\text{tot}}-C)f(C)}{Q(C)+P(C,C_{\text{tot}}-C)} = \frac{P(C_{\text{tot}}-C,C)f(C_{\text{tot}}-C)}{Q(C_{\text{tot}}-C)+P(C_{\text{tot}}-C,C)}. \tag{8.4.18}$$

A symmetry-breaking solution exists if equation (8.4.18) has solutions for which $C \neq C_{\text{tot}}/2$. Setting

$$G(C) = \frac{P(C,C_{\text{tot}}-C)f(C)}{Q(C)+P(C,C_{\text{tot}}-C)}, \tag{8.4.19}$$

we plot $G(C)$ and $G(C_{\text{tot}}-C)$ in Fig. 8.26a for various values of κ. For each κ, the intersections of the solid curve $G(C)$ with the dashed curve $G(C_{\text{tot}}-C)$ determines the possible steady-state values for C. As κ increases, the symmetric solution at $C = C_{\text{tot}}/2$ bifurcates into two additional non-symmetric solutions at a critical value κ_c, as illustrated in Fig. 8.26b. Note that $G(C)$ is a measure of the steady-state protein flux of U in the anterograde direction, which is balanced by the corresponding flux $G(C_{\text{tot}}-C)$ in the opposite direction. (In the given parameter regime, the steady-state fluxes of proteins U and V have the same magnitude). Fig. 8.26b shows that the stable flux is independent of κ for $\kappa < \kappa_c$ and a decreasing function of κ for $\kappa > \kappa_c$. A similar bifurcation scenario occurs when other parameters are varied such as the baseline flux rate γ.

8.5 Exercises

Problem 8.1 (Diffusion model of axonal length control). Consider a 1D continuum model of the diffusive transport of tubulin along an axon [431]. Let $c(x,t)$ denote the concentration of tubulin at position x along the axon at time t. Suppose that at time t the axon has length $L(t)$ so that $x \in [0,L(t)]$. The transport of tubulin is modeled macroscopically in terms of an advection–diffusion equation with an additional decay term representing degradation at a rate γ:

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} - V\frac{\partial c}{\partial x} - \gamma c.$$

Such a model can be derived from a more detailed stochastic model of active transport with V the effective drift due to motor-driven transport and D the effective diffusivity. It is assumed that there is a constant flux of newly synthesized tubulin from the cell body at $x = 0$ so that

$$-D\frac{\partial c}{\partial x}\bigg|_{x=0} = \sigma.$$

The flux at the growing end $x = L(t)$ is equal to the difference between the fluxes associated with microtubule assembly and disassembly:

$$-D\frac{\partial c}{\partial x}\bigg|_{x=L} = \varepsilon_l c(L) - \gamma_l,$$

where ε_l and γ_l are the rates of polymerization and depolymerization. Finally, the rate of growth is also taken to be proportional to the difference between these two fluxes according to

$$\frac{dL}{dt} = a\left[\varepsilon_l c(L(t),t)) - \gamma_l\right].$$

The constant a depends on the size of each tubulin dimer, the number of microtubules at the tip, and the cross-sectional area of the axon.

(a) Determine the steady-state solution $c(x)$ in terms of the steady-state length L using the boundary conditions at $x = L$. Hence, derive the following transcendental equation for L by imposing the boundary condition at $x = 0$:

$$F(L) \equiv \mathrm{e}^{-\lambda_- L} - \mathrm{e}^{-\lambda_+ L} = \frac{D\sigma}{\gamma}\frac{1}{c_L}(\lambda_+ - \lambda_-),$$

where $c_L = \varepsilon_L/\gamma_L$ and

$$\lambda_\pm = \frac{V}{2D}\left[1 \pm \sqrt{1 + 4D\gamma/V^2}\right].$$

Determine an approximation for L in (i) the small L regime and (ii) the large L and large V regime.

(b) Now suppose that diffusion is dominant, the concentration rather than the flux at $x = 0$ is fixed, and the rate of polymerization is infinitely fast. Then

$$\frac{\partial c}{\partial t} = \frac{\partial^2 c}{\partial x^2}, \quad 0 < x < L(t),$$

with

$$c(0,t) = c_0, \quad c(L(t),t) = 0,$$

and

$$-\frac{\partial c}{\partial x}\bigg|_{x=L(t)} = \beta\frac{dL(t)}{dt}.$$

We have set $D = 1$ for simplicity. Trying a solution of the form $c(x,t) = 1 + A\,\mathrm{erf}(x/2\sqrt{t})$, show that

$$u(x,t) = c_0 \left[1 - \frac{\mathrm{erf}[x/(2\sqrt{t})]}{\mathrm{erf}(\lambda)}\right],$$

where the error function $\mathrm{erf}(x) = (2/\sqrt{\pi}) \int_0^x \mathrm{e}^{-t^2} dt$, and λ satisfies the transcendental equation

$$\sqrt{\pi}\beta \mathrm{erf}(\lambda)\mathrm{e}^{\lambda^2} = 1.$$

This is known as the Neumann solution to the one-phase Stefan problem and has been used to model the growth of the acrosome of the sea cucumber *Thyone* during fertilization [505].

Problem 8.2 (Stochastic model of IFT). Consider a particle undergoing a unidirectional random walk along a single filament track as shown in Fig. 8.27. The track is modeled as a finite 1D lattice with lattice spacing ℓ. Suppose that at time t there are $N(t)+1$ lattice sites labeled $n = 0,\ldots,N$, with $n = 0$ corresponding to the $-$ end and $n = N(t)$ to the $+$ end. Suppose that $N(0) = 0$ and the particle starts at the minus end. During the jth cycle of the dynamics with $N(t) = j$, the particle walks from the minus to the plus end. The times τ between successive steps are taken to be independent, identically distributed random variables with a common waiting time density $\psi(\tau)$. When the particle reaches the current $+$ end, the length of the filament is increased by one lattice site to form the new $+$ end. Once the particle has reached this new lattice site, the hopping process reverses direction. After returning to the $-$ end the particle reverses direction again immediately, and the process continues iteratively. For simplicity the waiting time density is taken to be the same in both directions, and we are neglecting depolymerization at the plus end.

(a) Let $f_j(t)$ be the FPT density for the particle to travel from one end to the other when the length is $L_j = (j+1)\ell$. Similarly, let $g_n(t)$ be the probability density that the particle has just completed the nth visit to the $+$ end at time t and filament length has increased by one unit. Explain the meaning of the following iterative equations:

$$f_j(t) = \int_0^t \psi(\tau) f_{j-1}(t-\tau) d\tau,$$

with $f_0(t) = \psi(t)$ and

$$g_n(t) = \int_0^t \int_0^{t'} f_n(t-t') f_{n-1}(t'-t'') g_{n-1}(t'') dt'' dt'$$

for $n \geq 2$ with $g_1(t) = f_1(t)$.

(b) Using Laplace transforms show that for $n \geq 2$

$$\tilde{g}_n(s) = \tilde{f}_j(s) \prod_{k=1}^{n-1} \left[\tilde{f}_k(s)\right]^2$$
$$= \{\tilde{\psi}(s)\}^{(n+1)^2-2}.$$

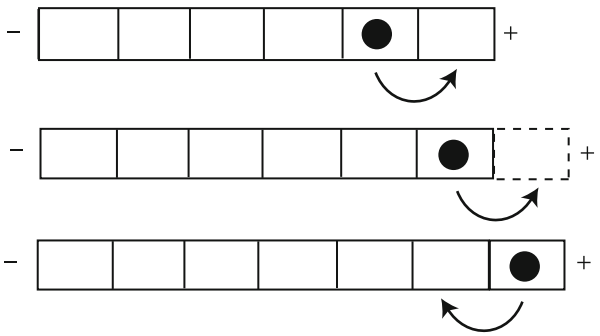


Fig. 8.27: Particle hopping along a one-dimensional filament that is modeled as a discrete lattice. When the particle reaches the + end of the filament, a lattice site is added to form the new + end and the particle reverses its direction. The particle also reverses direction at the − end

(c) Let $L(t)$ be the length of the flagellum at time t and introduce the length probability $P_j(t) = \text{Prob}[L(t) = L_j]$. Explain the iterative equation

$$P_j(t) = \int_0^t F_{j+1}(t-t')g_j(t')dt',$$

for $j \geq 1$, where $F_{j+1}(t-t')$ is the probability that the particle has not completed the $(j+1)$th trip to the plus end at time t, starting from the plus end at time t'. Show that

$$F_{j+1}(\tau) = \int_\tau^\infty \left[\int_0^{t'} f_{j+1}(t'-t'')f_j(t'')dt'' \right] dt'$$

for $j \geq 1$. Determine $\tilde{F}_j(s)$ and hence show that

$$\tilde{P}_j(s) = \frac{\widetilde{g_j}(s) - \tilde{g}_{j+1}(s)}{s},$$

with $g_0(t) = \delta(t)$. Determine $\sum_{j=0}^\infty \tilde{P}_j(s)$ and interpret the result.

(d) A useful way to characterize the stochastic growth of the filament is in terms of the mean and variance of the length $L(t)$. Let $\eta_n(t) = \sum_{j=0}^\infty j^n P_j(t)$ denote the nth moment of the distribution $P_j(t)$. Then

$$\langle L \rangle = 1 + \ell\eta_1(t), \ \langle \Delta L^2 \rangle = \ell^2[\eta_2(t) - \eta_1(t)^2].$$

Using part (c) show that

$$\tilde{\eta}_1(s) = \frac{1}{s}\sum_{j=1}^\infty \tilde{g}_j(s), \quad \tilde{\eta}_2(s) = \frac{1}{s}\sum_{j=1}^\infty (2j-1)\tilde{g}_j(s).$$

(e) Suppose that the waiting density has finite first and second moments so that for small s, $\tilde{\psi}(s) \sim 1 - \tau s$ and $\tilde{g}_j(s) \approx e^{-(j+1)^2 \tau s}$. The sums in part (d) can then be approximated for small s as

$$\tilde{\eta}_1(s) \sim \frac{1}{2s}\sqrt{\frac{\pi}{\tau s}}, \quad \tilde{\eta}_2(s) \sim \frac{1}{\tau s^2}.$$

Using a Tauberian theorem (see Box 7A), derive the large-t behavior

$$\eta_1(t) \sim \sqrt{t/\tau}, \quad \eta_2(t) \sim t/\tau.$$

Problem 8.3 (Search-and-capture model). Let $P(\tau)$ be the probability that a single MT finds a single kinetochore before time τ. From equation (8.2.1),

$$P(\tau) = p + \sum_{n=1}^{\infty} Q(\tau - \Delta\tau|n)P(n),$$

with $P(n) = p(1-p)^n$, $\Delta\tau = V_g/x$, and $Q(\tau|n)$ given by equation (8.2.4).

(a) Show that the probability of capture before time τ, $\tau > \Delta\tau$, is

$$P(\tau) = p + (1-p)(1 - \mathrm{e}^{-p(\tau-\Delta\tau)/T_{\text{cycle}}}).$$

(b) If $p \ll 1$ and $\Delta\tau \ll \tau$, then

$$P(\tau) \approx 1 - \mathrm{e}^{-p\tau/T_{\text{cycle}}}.$$

Using the expressions for p and T_{cycle} (see equations (8.2.2) and (8.2.3)), show that the optimal catastrophe frequency is $k_{\text{cat}} = V_g/x$.

(c) Now suppose that there are N independent searching MTs, and denote the probability that a single kinetochore is found before time τ by $P^{(N)}(\tau)$. Explain the formula

$$P^{(N)}(\tau) = 1 - (1 - P(\tau))^N.$$

Using the expression for $P(\tau)$ in part (b), show that the average time to capture is T_{capture}/N, where T_{capture} is the result for $N = 1$.

(d) Finally, consider N MTs and M kinetochores. Since the attachment of a kinetochore is independent of all other attachment events, the probability that all kinetochores will be attached to an MT before time τ is

$$P^{(N,M)}(\tau) = P^{(N)}(\tau)^M = (1 - \mathrm{e}^{-pN\tau/T_{\text{cycle}}})^M.$$

The corresponding density function is $f(t) = dP^{(N,M)}(t)/dt$. The most likely time t_0 when the last kinetochore is captured is obtained by finding the maximum of the function $f(t)$. Show that

$$t_0 = \frac{T_{\text{cycle}}}{pN} \ln M.$$

It turns out from numerical studies that t_0 is a reasonable estimate of the average time to capture all the kinetochores.

Problem 8.4 (Asymptotic analysis of continuum Hill model). Consider the steady-state continuum Hill model analyzed in Box 8B. We would like to calculate the mean velocity of the kMT given by $v = V(L/\eta)$, where $V(y)$ is the solution to the first-order ODE (in dimensionless units)

$$\frac{dV}{dy} = \delta(\alpha(y) - \beta(y))p(y), \quad V(0) = 0.$$

(a) Introduce the asymptotic expansions

$$V(y) = V_0(z) + \varepsilon V_1(z,\sigma) + O(\varepsilon^2),\ p(y) = p_0(z,\sigma) + \varepsilon p_1(z,\sigma) + O(\varepsilon^2)$$

with $V_1(z,\sigma)$ periodic in σ. Substituting into the ODE for $V(y)$ and collecting $O(1)$ terms, show that

$$\frac{dV_0(z)}{dz} + \frac{\partial V_1(z,\sigma)}{\partial \sigma} = \delta(\alpha(z) - \beta(z))p_0(z,\sigma),$$

with $p_0(z,\sigma)$ given by equations (8.2.16) and (8.2.19).

(b) Integrating the $O(1)$ equation over the interval $0 \le \sigma \le 1$ and exploiting the periodicity of V_1 obtain the following approximation for the mean velocity (expressed in terms of original variables),

$$v = \delta \int_0^L (\alpha(x) - \beta(x))A(x)I_0(f(x)/k_BT)dx + O(\varepsilon),$$

where

$$A(x) = C\exp\left(\frac{r(x)}{k_BT} - \frac{\delta\gamma}{k_BT}\int^x I_0^2(f(x')/k_BT)[\beta(x') - \alpha(x')]dx'\right).$$

(c) Why is the expression for v in part (b) preferable for numerical calculations rather than the one obtained by substituting equation (8.2.20) for $p_0(x)$ directly into

$$v = \delta \int_0^L (\alpha(x) - \beta(x))p_0(x)dx + O(\varepsilon)?$$

Problem 8.5 (Spindle length control). Consider the model of spindle length control during *Drosophila* metaphase given by equations (8.2.33).

(a) Derive the steady-state equations (8.2.34) for the spindle length S and length of overlap L between inter polar MTs.

(b) Determine conditions for the stability of the steady-state.

(c) Consider the following basal parameter values: $N = 100$, $S_0 = 10\,\mu$m, $\delta = 4$ nm, $V_{\text{slide}} = 0.1\,V_0 = 0.01$, $V_1 = 0.06$, $V_{\text{poly}} = 0.06$, $\alpha = 100$, $\beta = 80$, $F_{\text{Kt}} = 100$, and $k_B T = 4.2 \times 10^{-3}$. All velocities are $\mu m/s$ and all forces are $pN/\mu m$. Plot the steady state lengths S and L as a function of α over the range $100 \leq \alpha \leq 10^4$. Comment on what happens if α becomes too small.

Chapter 9
Self-Organization in Cells II: Reaction-Diffusion Models

In the previous chapter, we focused on the role of active processes such as polymerization on the self-organization of cytoskeletal structures. We now turn to another major mechanism for self-organization within cells (and between cells), namely, the interplay between diffusion (Chap. 2) and nonlinear chemical reactions (Chap. 6). Historically speaking, the idea that a reaction–diffusion system can spontaneously generate spatiotemporal patterns was first introduced by Turing in his seminal 1952 paper [646]. Turing considered the general problem of how organisms develop their structures during the growth from embryos to adults. He established the principle that two nonlinearly interacting chemical species differing significantly in their rates of diffusion can amplify spatially periodic fluctuations in their concentrations, resulting in the formation of a stable periodic pattern. The Turing mechanism for morphogenesis was subsequently refined by Gierer and Meinhardt [216], who showed that one way to generate a Turing instability is to have an antagonistic pair of molecular species, a slowly diffusing chemical activator and a quickly diffusing chemical inhibitor, which they applied to a number of specific biological systems. Over the years, the range of models and applications of the Turing mechanism expanded dramatically [461],[1] in spite of the fact that most experimental findings suggested that morphogenesis was often guided by explicit spatial cues, based on the localization of specific proteins or RNA [687]. Indeed, for many years the only direct experimental evidence for spatiotemporal patterning of molecular concentrations came from the inorganic Belousov–Zhabotinsky reaction [700], until Kondo and Asai demonstrated the occurrence of the Turing mechanism in studies of animal coat patterning [350]. Recent advances in live cell imaging and gene knockout protocols are now allowing for a closer connection between theories of pattern formation and cell biology. Moreover, there are a growing number of examples of spatiotemporal patterns at the intracellular level, which will be the focus of this chapter.

[1] Biological pattern formation has also been studied within the context of ecology [385, 539] and systems neuroscience [75, 172]; in the latter case nonlocal synaptic interactions drive pattern forming instabilities rather than diffusion.

P.C. Bressloff, *Stochastic Processes in Cell Biology*, Interdisciplinary Applied Mathematics 41, DOI 10.1007/978-3-319-08488-6_9

We begin by considering the formation of intracellular protein gradients, which are characterized by spatial variations in the activity state (eg., phosphorylation state) of a protein due to some localized source of activation or deactivation (Sect. 9.1). We address the robustness of intracellular gradients in the presence of intrinsic noise due to low protein copy numbers. In Sect. 9.2 we review the theory of Turing pattern formation in terms of deterministic reaction–diffusion (RD) equations. The effects of intrinsic noise on spontaneous pattern formation is considered in Sect. 9.3, where we construct an RD master equation. The latter is obtained by discretizing space and treating spatially discrete diffusion as a hopping reaction. Carrying out a linear noise approximation of the master equation leads to an effective Langevin equation, whose power spectrum provides a means of extending the definition of a Turing instability to stochastic systems, namely, in terms of the existence of a peak in the power spectrum at a nonzero spatial frequency. One thus finds that noise can significantly extend the range over which spontaneous patterns occur. Finally, we consider two examples of spontaneous self-organization in cells: the Min system of proteins responsible for regulating the initiation site of bacterial cell division (Sect. 9.4) and cell polarization (Sect. 9.5).

In keeping with previous chapters, we will concentrate on relatively simple, analytically tractable models of reaction–diffusion systems. A complementary approach involves developing more biologically realistic multi-scale computational models, which include details of the structure of individual macromolecules, the biochemical network of signaling pathways, the aqueous environment of the cytoplasm, the mechanical properties of the cytoskeleton, and the geometry of the cell [43, 171, 192, 611]. A major challenge in the computational modeling of reaction–diffusion systems is how to efficiently couple stochastic chemical reactions involving low copy numbers with diffusion in complex environments. One approach is to consider a spatial extension of the Gillespie algorithm for well-mixed chemical reactions [217, 218] (Sect. 6.8) using a mesoscopic compartment-based method, although there are subtle issues with regard to choosing the appropriate compartment size [283, 292, 293, 647]. Alternatively, one can combine a coarse-grained deterministic reaction–diffusion model in the bulk of the domain with individual particle-based Brownian dynamics in certain restricted regions [10, 168, 169, 193]; in this case considerable care must be taken in the choice of boundary conditions at the interface between the two domains.

9.1 Intracellular Protein Concentration Gradients

It has been known for some time that concentration gradients play a crucial role in the spatial regulation of patterning during development [370, 591, 672, 686, 687]. That is, a spatially varying concentration of a morphogen protein drives a corresponding spatial variation in gene expression through some form of concentration thresholding mechanism. For example, in regions where the morphogen concentration exceeds a particular threshold, a specific gene is activated. Hence, a continuously varying morphogen concentration can be converted into a discrete spatial pattern of differentiated gene expression across a cell population. The basic mechanism

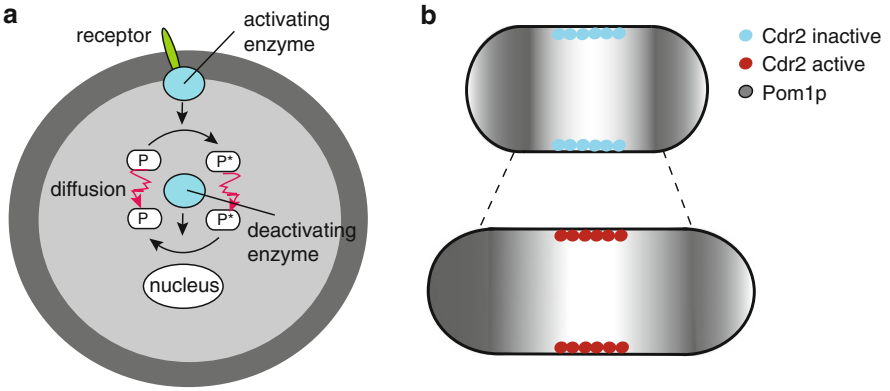


Fig. 9.1: (**a**) Cartoon of a protein modification cycle, in which an inactive form P is converted to an active form P^* at the plasma membrane. Both forms diffuse in the cytoplasm, resulting in deactivation of P^* by cytoplasmic enzymes. (**b**) Pom1 concentration gradient in fission yeast is highest at the poles and lowest in the midcell region where Cdr2 concentrates in cortical nodes. In early interphase (short cells), Pom1 at the midcell is present at a sufficient concentration to inhibit Cdr2 and the transition to mitosis. As cells grow, the midcell Pom1 concentration decreases until it crosses a threshold that relieves Cdr2 inhibition thereby promoting mitosis

of morphogen gradient formation is thought to involve a localized source of protein production within the embryo, combined with diffusion away from the source and subsequent degradation.

There is emerging experimental evidence that concentration gradients not only arise within the context of embryonic development, but are also found within individual cells, typically taking the form of a concentration gradient in some active protein [276, 331]. An important difference between intracellular gradients and multicellular morphogen gradients is that degradation does not play a significant role in the formation of intracellular gradients. This is a consequence of the fact that the lifetime of a typical protein exceeds the duration of the cellular process regulated by the presence of a gradient. Instead, some modification in the protein, such as its phosphorylation state, changes as it moves away from the catalytic source of the modification. The existence of an intracellular gradient was first predicted theoretically by Brown and Kholodenko [78] and has subsequently been found to play a role in a wide range of cellular processes, including cell division, polarity, and mitotic spindle dynamics:

1. An important component of many signal transduction pathways is the reversible cycling between an inactive and an active protein state, which is catalyzed by opposing activator and deactivator enzymes. For example, a kinase and phosphatase acting on phosphoproteins, or a GEF (guanine-nucleotide-exchange factor) and GAP (GTPase-activating protein) acting on small G-proteins such as proteins of the Ras and Rho families. A concentration gradient in these signaling cycles can then be generated by the spatial segregation of the opposing enzymes [331, 459]. One such mechanism is the phosphorylation of proteins by a

membrane-bound kinase, which are then dephosphorylated by a cytosolic phosphatase (see Fig. 9.1a). This results in a gradient of the phosphorylated protein, with a high concentration close to the cell membrane and a low concentration within the interior of the cell. As the cell grows in size, the surface-to-volume ratio decreases and membrane-activated proteins have to diffuse over longer distances in order to reach their target such as the nucleus. Hence, the proteins become progressively deactivated towards the cell interior, thus providing a mechanism for coupling cell growth with the cell cycle. Indeed, activated spatial gradients have been observed during cell mitosis [98, 315, 478] involving for example the small GTPase Ran, which assists in the formation of the mitotic spindle by biasing microtubule growth towards the chromosomes, (see also Sect. 8.2). Activity gradients may also play a role in the localization of the RhoGTPase Cell division control protein 42, which is a regulator of actin polymerization, resulting in a coupling between cell shape and protein activation [439].

2. One of the best studied intracellular gradients involves the *dual-specificity tyrosine phosphorylation-regulated kinase* (DYRK) Pom1p in fission yeast. Pom1p forms a concentration gradient within the rod-shaped yeast cell, with the highest concentrations at the cell tips and the lowest concentrations at the cell center. This is achieved by a combination of localized binding to the membrane at the cell tips, diffusive spreading within the membrane, followed by membrane unbinding and diffusion in the cytoplasm until rebinding at the tips. It is thought that Pom1p inhibits cell division, acting to localize the cell division factor Mid1 towards the cell mid-plane [99, 490]. Pom1p also phosphorylates and suppresses the activity of Cdr2p, which is a promoter of cell mitosis that localizes to nodes within the membrane at the center [421, 455]. In a short cell, the concentration of Pom1p at the cell center is relatively high so that Cdr2 activity is inhibited. However, as the cell grows, the midcell concentration of Pom1p decreases resulting in activation of Cdr2 and initiation of mitosis (see Fig. 9.1b). Thus, the formation of a spatial concentration gradient again provides a mechanism for coupling cell growth with the cell cycle.
3. Spatial concentration gradients have also been found to play a role in cell division in small bacterial cells with spatial extents of only a few microns [541, 623]. One well-studied example is the oscillatory dynamics of Min proteins in *E. coli*, which creates a time-averaged concentration gradient that directs localization of the cell division machinery [277, 284, 356, 396, 434]. Such an oscillation is the result of a dynamic instability of the underlying reaction–diffusion system (see Sect. 9.4).

9.1.1 Spatially Distributed Signaling Cascades

We begin by considering the simplest system capable of generating a stationary concentration gradient, consisting of an activating enzyme located in the cell membrane and a deactivating enzyme freely diffusible in the cytoplasm [78, 331] (see

Fig. 9.1a). For simplicity, consider an effective 1D geometry representing, for example, a cylindrical bacterial cell of length L. Suppose that a kinase is localized to a pole at $x = 0$, which generates a flux of phosphorylated protein at a rate v_+, whereas the phosphatase is distributed throughout the cytoplasm and deactivates each protein at a rate v_-. Assume that the phosphatase is far from saturation so that $v_- = k_- c^*$, where c^* is the active protein concentration. It follows that $c^*(x,t)$ evolves according to the diffusion equation

$$\frac{\partial c^*}{\partial t} = D\frac{\partial^2 c^*}{\partial x^2} - k_- c^*(x,t), \tag{9.1.1}$$

supplemented by the boundary conditions

$$-D\frac{\partial c^*}{\partial x}\bigg|_{x=0} = v_+, \quad \frac{\partial c^*}{\partial x}\bigg|_{x=L} = 0. \tag{9.1.2}$$

This has the steady-state solution

$$c^*(x) = c^*(0)\left(\frac{\mathrm{e}^{x/\lambda} + \mathrm{e}^{2L/\lambda}\mathrm{e}^{-x/\lambda}}{1+\mathrm{e}^{2L/\lambda}}\right), \quad \lambda = \sqrt{\frac{D}{k_-}}. \tag{9.1.3}$$

The concentration $c^*(0)$ can be determined from the boundary condition at $x = 0$. Note that when $L \ll \lambda$, the gradient is approximately linear, whereas when $\lambda L \gg \lambda$ it decays exponentially with length constant λ.

For biophysically reasonable parameter values, one typically finds that the protein gradient is very steep, which means that it is unlikely to reach the nucleus of large cells such as developing neurons. Several mechanisms have been suggested that could produce longer-range signal transduction, including spatially distributed signaling cascades, active transport, and traveling waves of protein phosphorylation [459]. Here we will consider the particular example of signaling cascades as developed in [460]. That is, suppose there exists a cascade of protein modification cycles, in which each cycle involves transitions between inactive and active forms of a signaling protein. At each level of the cascade, the active form of the protein catalyzes the activation of the protein at the next downstream level (see Fig. 9.2a). Let $c_n^*(x,t)$ and $c_n(x,t)$ denote the concentration of activated and deactivated protein, respectively, at the nth level of the cascade. Assume that the total concentration of protein at each cascade level is fixed at c_n^{tot} so that $c_n(x,t) = c_n^{\mathrm{tot}} - c_n^*(x,t)$. We have

$$\frac{\partial c_1^*}{\partial t} = D\frac{\partial^2 c_1^*}{\partial x^2} - v_1^-(x,t), \tag{9.1.4a}$$

$$\frac{\partial c_n^*}{\partial t} = D\frac{\partial^2 c_n^*}{\partial x^2} + v_n^+(x,t) - v_n^-(x,t), \quad n = 2,\ldots,N, \tag{9.1.4b}$$

where v_n^+ and v_n^- are the phosphorylation and dephosphorylation rates, respectively. Equation (9.1.4) are supplemented by the boundary conditions

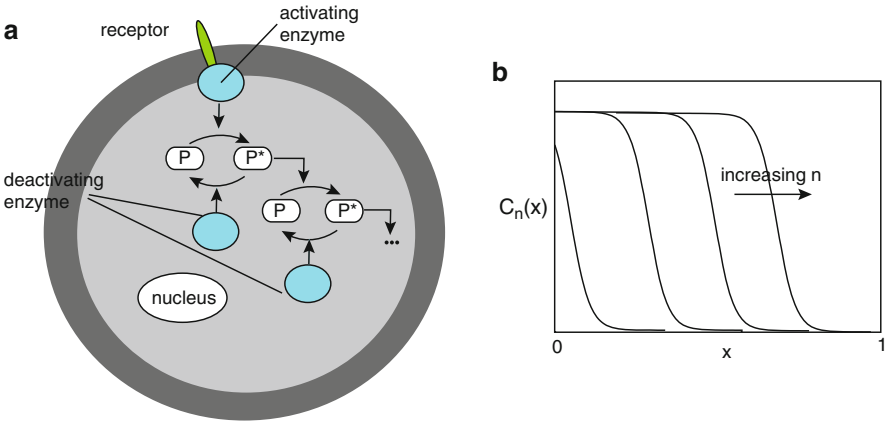


Fig. 9.2: Spatial propagation of activated forms for a cascade of protein modification cycles. (**a**) Phosphorylation cycles within a cell. (**b**) Schematic illustration of steady-state concentration profiles at successive levels n of the cascade for $\gamma < 1$ in nondimensionalized version of the model

$$-D\frac{\partial c_1^*}{\partial x}\bigg|_{x=0} = v_1^+, \quad \frac{\partial c_1^*}{\partial x}\bigg|_{x=L} = 0 \tag{9.1.5a}$$

$$\frac{\partial c_n^*}{\partial x}\bigg|_{x=0} = 0 = \frac{\partial c_n^*}{\partial x}\bigg|_{x=L}, \quad n = 2,\ldots,N. \tag{9.1.5b}$$

Assuming Michaelis–Menten kinetics for the enzymatic reactions (see Box 6B), we have

$$v_n^- = V_n^- \frac{c_n^*(x,t)}{K_n^- + c_n^*(x,t)}, \quad n = 1,\ldots,N \tag{9.1.6a}$$

$$v_1^+ = V_1^+ \frac{c_1^{\mathrm{tot}} - c_1^*(0,t)}{K_1^+ + c_1^{\mathrm{tot}} - c_1^*(0,t)}, \tag{9.1.6b}$$

$$v_n^+ = \hat{V}_n^+ c_{n-1}^*(x,t) \frac{c_n^{\mathrm{tot}} - c_n^*(x,t)}{K_n^+ + c_n^{\mathrm{tot}} - c_n^*(x,t)}, \quad n = 2,\ldots,N. \tag{9.1.6c}$$

Suppose, for simplicity, that all parameters are independent of the cascade level n: $K_n^- = K_-, V_n^- = V_-$, $c_n^{\mathrm{tot}} = c_{\mathrm{tot}}$ for $n = 1,\ldots,N$ and $K_n^+ = K_+, \hat{V}_n^+ c_{\mathrm{tot}} = V_+$ for $n = 2,\ldots N$. We also assume that the Michaelis–Menten kinetics operate in the linear regime (small concentrations). Nondimensionalizing the equations by setting

$$C_n = c_n^*/c_{\mathrm{tot}}, \quad x' = \sqrt{\frac{k_-}{D}}x, \quad t' = k_+ t, \quad \gamma = \frac{k_-}{k_+},$$

with $k_\pm = V_\pm/K_\pm$, and dropping primes gives

$$\frac{\partial C_1}{\partial t} = \gamma \frac{\partial^2 C_1}{\partial x^2} - \gamma C_1, \tag{9.1.7a}$$

$$\frac{\partial C_n}{\partial t} = \gamma \frac{\partial^2 C_n}{\partial x^2} - \gamma C_n + (1 - C_n) C_{n-1}, \quad n = 2, \ldots, N, \tag{9.1.7b}$$

$$-\frac{\partial C_1}{\partial x}\bigg|_{x=0} = \nu (1 - C_1)_{x=0}, \quad D \frac{\partial C_1}{\partial x}\bigg|_{x=L} = 0 \tag{9.1.7c}$$

$$\frac{\partial C_n}{\partial x}\bigg|_{x=0} = 0 = \frac{\partial C_n}{\partial x}\bigg|_{x=L}, \quad n = 2, \ldots, N, \tag{9.1.7d}$$

with

$$\nu = \frac{V_1^+}{K_1^+} \sqrt{\frac{1}{D k_-}}.$$

The crucial parameter that determines the degree of spread of activity is γ, which is the ratio of the deactivation and activation rates. Numerically solving these equations shows that for $\gamma < 1$, an initial activation signal at the boundary $x = 0$ propagates into the domain, converging to a steady-state solution consisting of stationary front-like profiles that are shifted further into the domain at higher levels of the cascade [460]. This is illustrated schematically in Fig. 9.2b. On the other hand, if $\gamma > 1$, then the activated proteins fail to propagate into the domain, and the concentrations of activated proteins decay rapidly close to the plasma membrane. In this latter regime, it is possible to obtain analytical approximations for the concentration profiles. First note that the steady-state equation for C_1 can be solved as in the single-level model. In particular, for a large domain, it has the exponential form $C_1(x) = C_1(0)\mathrm{e}^{-x}$. The steady-state solution at successive levels can then be determined from the approximate recurrence relation [460]

$$\gamma \frac{d^2 C_n}{dx^2} - \gamma C_n + C_{n-1} = 0, \tag{9.1.8}$$

assuming $C_n \ll 1$. When $\gamma > 1$, one can solve for $C_n(x)$ using a polynomial expansion of the form

$$C_n(x) = \left(\sum_{m=0}^{n-1} Q_n^{(m)} x^m \right) \mathrm{e}^{-x}.$$

In the more interesting regime $\gamma < 1$, the polynomial construction breaks down near the boundary $x = 0$. Nevertheless, one can still determine the leading order behavior in the tail of the profiles, $x \gg 1$, for which

$$C_n(x) \approx \frac{C_1(0)}{(n-1)!(2\gamma)^{n-1}} x^{n-1} \mathrm{e}^{-x}.$$

This can then be used to determine the spread of activated protein at the nth level [460]. To a first approximation, the front profile shifts an amount (in physical units)

$$\Delta X = (1-\gamma)\ln(1/\gamma)\sqrt{k_-/D} \tag{9.1.9}$$

from one cascade level to the next. Qualitatively similar behavior is found when the full Michaelis–Menten kinetics is used, except now the condition for propagation failure depends on γ and the degree of saturation.

9.1.2 Robustness of Concentration Gradients

Intracellular protein gradients provide a mechanism for determining spatial position within a cell so that, for example, cell division occurs at the appropriate time and location. Similarly, developmental morphogen gradients control patterns of gene expression so that each stage of cell differentiation occurs at the correct spatial location within an embryo. For gradient mechanisms to be biologically effective, however, position determination has to be robust to both intrinsic and extrinsic noise fluctuations. Recall from Sect. 6.1 that extrinsic noise is usually associated with cell-to-cell variations in environmental factors, whereas intrinsic noise refers to fluctuations within a cell due to biochemical reactions involving small numbers of molecules. Both forms of noise can affect positional accuracy as illustrated in Fig. 9.3. Here we will explore the effects of intrinsic noise on the simple gradient producing systems shown in Fig. 9.1, following the analysis of Tostevin et al. [637, 639]. For simplicity, consider a cylindrical geometry with the dimension d of the system given by $d = 2$ if the gradient is restricted to the membrane (as in Pom1p) or $d = 3$ if it is in the cytoplasm. Take the x-axis to be the axial coordinate and assume that the concentration is uniform in the transverse coordinates(s). Take the length of the cell to be L with $0 < x < L$ and assume that there is a source at $x = 0$ where proteins are produced at a rate J per unit area or circumference. The concentration gradient is given by Eq. (9.1.3), written in the form

$$c(x) = \frac{J\lambda}{D}\mathrm{e}^{-x/\lambda}. \tag{9.1.10}$$

Suppose that the concentration gradient has to identify a particular spatial location along its length. This location could be determined by the point x_T at which the monotonically decreasing concentration profile crosses a threshold c_T. The system will then be divided into two domains: an active domain $0 \leq x < x_T$ for which $c(x) > c_T$ and an inactive domain $x_T \leq x \leq L$ where $c(x) \leq c_T$. Noise in the local protein concentration will cause fluctuations in the threshold position. Tostevin et al. [639] estimated the uncertainty in the position due to intrinsic noise. Recall from Sect. 6.2

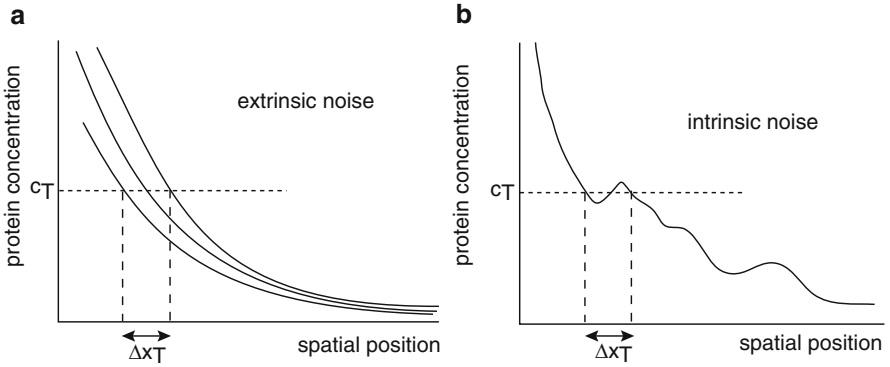


Fig. 9.3: Effect of noise on the positional information provided by concentration gradients. (**a**) Extrinsic noise in the protein gradient leads to a varying profile from one cell to another. Variation in the position at which the gradient concentration drops below a critical level c_T leads to imprecision in the specification of position x_T. (**b**) Intrinsic noise within a single gradient also leads to imprecise positional information. Redrawn from Howard [276]

that a simple reaction involving protein production and degradation exhibit, Poisson statistics. In particular, if $n(\mathbf{x})$ is the random number of molecules in a volume a^d centered at $\mathbf{x} \in \mathbb{R}^d$, then

$$\langle n(\mathbf{x})^2 \rangle - \langle n(\mathbf{x}) \rangle^2 = \langle n(\mathbf{x}) \rangle.$$

Dividing through by the volume, we find that the fluctuations in the concentration are

$$\langle c(\mathbf{x})^2 \rangle - \langle c(\mathbf{x}) \rangle^2 = \frac{\langle c(\mathbf{x}) \rangle}{a^d},$$

with $\langle c(\mathbf{x}) \rangle$ given by Eq. (9.1.10). We identify a with the size of the region measuring the concentration, which could be the size of a receptor with which the gradient proteins interact.

The uncertainty Δx in spatial location can now be estimated using

$$\Delta x |c'(x_T)| = \sqrt{\mathrm{Var}\, c(\mathbf{x})} = \sqrt{\frac{c(x_T)}{a^d}}, \tag{9.1.11}$$

which implies that

$$\Delta x = \sqrt{\frac{\lambda D}{J a^d}} \mathrm{e}^{x_T/2\lambda}. \tag{9.1.12}$$

The optimal decay rate λ then depends on whether the flux J is kept constant or the number of molecules N is kept constant. In the former case, minimizing Δx with respect to λ shows that the optimal value is $\lambda = x_T/2$. In the latter case, the flux J can be expressed in terms of the total number of proteins N according to (for $d = 2$)

$$N = L_\perp \int_0^L \rho(x)dx = \frac{JL_\perp \lambda^2}{D}\left[1 - e^{-L/\lambda}\right] \approx \frac{JL_\perp \lambda^2}{D},$$

assuming $\lambda \ll L$ and $L_\perp$ is the circumference of the cell. This implies that $J = ND/(L_\perp \lambda^2)$ and thus

$$\frac{\Delta x}{L} = \frac{1}{L}\sqrt{\frac{\lambda^3 L_\perp}{Na^2}} e^{x_T/2\lambda}.$$

This formula implies that the uncertainty in position decreases with copy number according to $N^{-1/2}$ and is a nonlinear function of λ with a global minimum given by

$$\frac{d\Delta x(\lambda)}{d\lambda} = \frac{3}{2\lambda}\Delta x - \frac{x_T}{2\lambda^2}\Delta x,$$

which yields $\lambda_{\min} = x_T/3$.

Typical parameter values for membrane gradients in bacteria such as fission yeast are cell length $L = 10\,\mu\text{m}$, circumference $L_\perp = 6\,\mu\text{m}$, diffusivity $D = 0.1\,\mu\text{m}^2\text{s}^{-1}$, decay length $\lambda = 2\,\mu\text{m}$, and detector size $a = 0.01\,\mu\text{m}$. Taking $x_T = 4\,\mu\text{m}$ and a reasonable copy number $N = 4{,}000$ leads to the following estimate of the uncertainty in position: $\Delta x \approx 2L$. This implies that the cell must carry out some additional processing at the signal detection level in order to reduce the uncertainty in position. One mechanism for achieving high precision even for low copy numbers is time averaging [237, 637, 639]. (Alternative mechanisms of noise reduction are considered by [276, 278, 559] for intracellular gradients and [57, 161, 167] for developmental morphogen gradients.) Suppose that a receptor, for example, integrates the concentration of the gradient protein over a time interval of length τ. The detector is then able to perform $N_\tau = \tau/\tau_D$ independent measurements of the concentration, where τ_D is the time for correlations in the concentration to decay. A rough estimate of τ_D (ignoring logarithmic corrections) is $\tau_D = a^2/D \approx 10^{-3}$. From the law of large numbers, we expect the uncertainty in position after time averaging becomes

$$\overline{\Delta x} = \frac{\Delta x}{\sqrt{N_\tau}} = \sqrt{\frac{\tau_D}{\tau}}\Delta x \approx \frac{0.1}{\sqrt{\tau}}L.$$

It follows that an integration time of 100 s would lead to uncertainty in position that is only 1 % of the cell length.

In the above analysis we considered the accuracy of positional information based on a single concentration gradient. Another example of a non-uniform distribution of proteins, such as Pom1 in fission yeast, involves two opposing gradients that are used to determine the position of the center of the cell in preparation for cell division [99, 421, 455, 490]. The effects of intrinsic noise on the precision of specifying the mid-point of a fission yeast cell has been analyzed by Tostevin et al. [638, 639], along similar lines to the single gradient case. The basic model is illustrated in Fig. 9.4. It is assumed that Pom1 associates with the membrane at the cell poles, resulting in an effective polar flux J at each end. This naturally generates a symmetric concentration profile with a minimum at the center through a combination of lateral membrane diffusion and dissociation from the membrane. Assume that

the membrane concentration around the circumference of the cell is uniform, and let $c(x)$ be the density of proteins (per unit area) as a function of the axial coordinate x, $x \in [-L/2, L/2]$, where L is cell length. Then $c(x)$ evolves according to the reaction–diffusion equation

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} - \mu c, \tag{9.1.13}$$

where μ is the rate of membrane dissociation, supplemented by the boundary conditions

$$D\frac{\partial c}{\partial x}\bigg|_{x=\pm L/2} = \pm J. \tag{9.1.14}$$

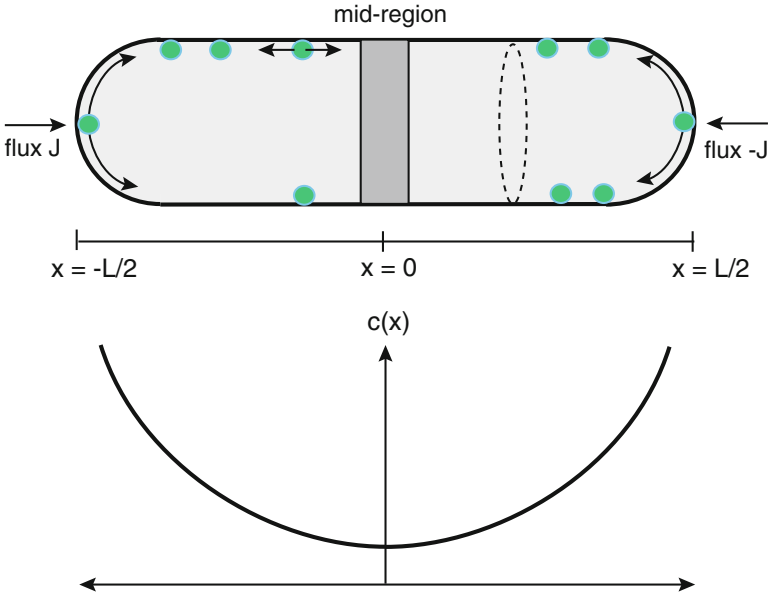


Fig. 9.4: A schematic illustration of a model for the formation of two opposing protein gradients within the membrane of a rod-like cell such as fission yeast [638]. Diffusing proteins within the cytoplasm bind to the membrane at the poles and then undergo lateral diffusion within the membrane. When this is combined with dissociation from the membrane, a symmetric protein concentration profile is set up with a minimum at the mid-plane. The cytoplasmic diffusion and membrane binding are represented as polar fluxes

Solving the boundary value problem (see Ex. 9.1) yields

$$c(x) = \frac{J}{D}\frac{\lambda \cosh(x/\lambda)}{\sinh(L/2\lambda)}, \tag{9.1.15}$$

where $\lambda = \sqrt{D/\mu}$ is the characteristic decay length of the spatial gradient. Clearly the deterministic concentration has a minimum at the center $x = 0$ with

$$c(0) = \frac{J}{D}\frac{\lambda}{\sinh(L/2\lambda)}. \tag{9.1.16}$$

Now suppose that intrinsic noise generates an uncertainty in the concentration at the center given by $\Delta c = c(0)/a^2$ with a the size of the protein detector. Since the deterministic concentration has a minimum at $x = 0$, it is necessary to Taylor expand to second order, in order to determine the uncertainty Δx in position, that is,

$$\Delta c = \frac{1}{2}|c''(0)|\Delta x^2. \tag{9.1.17}$$

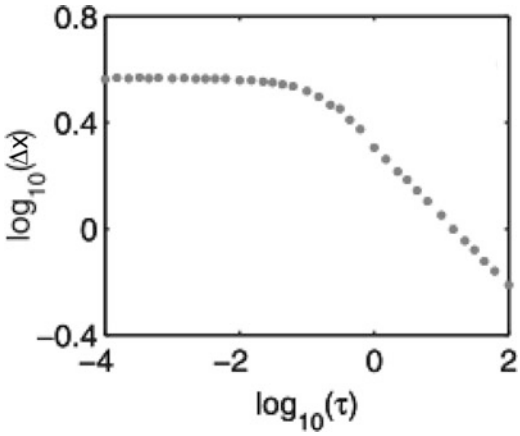


Fig. 9.5: Variation of uncertainty in position Δx with the integration time τ based on numerical simulations of a stochastic version of the dual gradient model. Parameter values are $L = 10\,\mu$m, $\mu = 0.36\,\text{s}^{-1}$, $J = 6\,\mu\text{m}^{-1}\text{s}^{-1}$, $D = \mu\text{m}^2\text{s}^{-1}$, and $a = -.01\,\mu$m (Adapted from Tostevin et al. [639])

One thus finds that (see Ex. 9.1)

$$\Delta x = \left(\frac{4D\lambda^3\sinh(L/2\lambda)}{Ja^2}\right)^{1/4}. \tag{9.1.18}$$

As in the case of a single concentration gradient, much greater precision can be obtained by time averaging. However, greater care has to be taken with regard temporal correlations when estimating the minimum time τ_D required for independent measurements in two-dimensional domains such as the cell membrane (see [639] for details). An example plot of how the uncertainty Δx varies with integration time

τ is shown in Fig. 9.5, based on stochastic simulations of a finite number of proteins. The main observation is that there are two distinct regimes, with Δx approximately independent of τ for small τ and $\Delta x \sim \tau^{-1/4}$ for large τ as predicted from the theory.

9.2 Theory of Turing Pattern Formation

One of the necessary conditions for the formation of a concentration gradient is that there exists some local source of intracellular proteins or a preexisting spatial variation of mRNA in an embryo. In other words, an underlying spatial symmetry of the cell is broken explicitly. An alternative mechanism for the formation of spatial patterns, via spontaneous symmetry breaking, is the diffusion-driven instability first hypothesized by Turing [646]. Typically, such an instability is modeled in terms of a system of reaction–diffusion (RD) equations. These describe two or more chemical species interacting via short-range activation kinetics (slow diffusion) and long-range inhibition kinetics (fast diffusion). In the absence of diffusion, the system converges to a homogeneous stable steady state, whereas the addition of diffusion can destabilize the homogeneous state resulting in a spatially varying pattern (provided that the physical domain is sufficiently large). One major biological application of the Turing mechanism has been in development, based on the reaction–diffusion of morphogens [461, 488]. More recently, however, the Turing mechanism has been used to account for various spatiotemporal patterns at the intracellular level, including oscillations of *Min protein* concentrations in *E. coli* [396] (Sect. 9.4) and the spatial variation of membrane-bound proteins in polarized cells [231, 437] (Sect. 9.5). Before considering explicit applications in cell biology, we review the basic theory of Turing pattern formation, covering the deterministic case here and the stochastic case in Sect. 9.3.

For concreteness, we will focus on a two-component system in two spatial dimensions, consisting of chemical concentrations $u(\mathbf{x},t)$ and $v(\mathbf{x},t)$ with $\mathbf{x} \in \mathbb{R}^2, t \in \mathbb{R}^+$. We will assume that the system is restricted to a bounded square domain Ω of size L so that $0 \le x \le L$ and $0 \le y \le L$. The standard RD model takes the form

$$\frac{\partial u}{\partial t} = D_u \nabla^2 u + f(u,v) \tag{9.2.1a}$$

$$\frac{\partial u}{\partial t} = D_v \nabla^2 u + g(u,v). \tag{9.2.1b}$$

Here D_u and D_v are the corresponding diffusion coefficients, and the nonlinear functions f, g describe the chemical reactions. The RD system is typically supplemented by a no-flux boundary condition

$$\mathbf{n} \cdot \nabla u = 0, \quad \mathbf{n} \cdot \nabla v = 0, \quad \text{for } \mathbf{x} \in \partial\Omega,$$

where $\mathbf{n}$ is the outward normal on the boundary $\partial\Omega$. Suppose that there exists a homogeneous stationary state (u^*, v^*) for which $f(u^*, v^*) = g(u^*, v^*) = 0$. The basic idea of the Turing mechanism is that the stationary state is stable to perturbations in the absence of diffusion, but is unstable to spatially inhomogeneous perturbations when diffusion is present. Linear stability analysis can be used to identify the fastest growing perturbations, which can be expressed in terms of linear combinations of eigenmodes of the associated linear operator. However, in order to determine whether or not these growing patterns themselves stabilize, it is necessary to go beyond linear theory by using perturbation methods to derive a system of nonlinear ODEs for the amplitude of a given pattern.

9.2.1 Linear Stability Analysis

Linearizing equations (9.2.1) about the homogeneous state (u^*, v^*) by setting $U = u - u^*, V = v - v^*$, leads to the linear system

$$\frac{\partial}{\partial t}\begin{pmatrix} U \\ V \end{pmatrix} = \mathbb{L}\begin{pmatrix} U \\ V \end{pmatrix} \equiv \begin{pmatrix} f_u & f_v \\ g_u & g_v \end{pmatrix}_{u^*,v^*}\begin{pmatrix} U \\ V \end{pmatrix} + \begin{pmatrix} D_u & 0 \\ 0 & D_v \end{pmatrix}\begin{pmatrix} \nabla^2 U \\ \nabla^2 V \end{pmatrix} \tag{9.2.2a}$$

supplemented by no-flux boundary conditions. Here $f_u = \partial f/\partial u$. Setting $\mathbf{U} = (U, V)^T$, the general solution of Eq. (9.2.2) can be written as

$$\mathbf{U}(\mathbf{x}, t) = \sum_{\mathbf{k}} \mathbf{c}_k e^{i\mathbf{k}\cdot\mathbf{x}} e^{\lambda(k)t},$$

where $\mathbf{c}_k e^{i\mathbf{k}\cdot\mathbf{x}}$ and $\lambda(k)$ form an eigenvalue pair of the linear operator $\mathbb{L}$, parameterized by the wavenumber $k = |\mathbf{k}|$. The k-dependence of the eigenvalue $\lambda(k)$ is known as a dispersion relation. Substitution of the general solution into Eq. (9.2.2) yields a characteristic equation for $\lambda(k)$:

$$\left|\mathbf{A} - \mathbf{D}k^2 - \lambda(k)\mathbf{I}\right| = 0, \tag{9.2.3}$$

with

$$\mathbf{A} = \begin{pmatrix} f_u & f_v \\ g_u & g_v \end{pmatrix}_{u^*,v^*}, \quad \mathbf{D} = \begin{pmatrix} D_u & 0 \\ 0 & D_v \end{pmatrix}.$$

Evaluating the determinant, we obtain a quadratic equation for λ:

$$\lambda^2 + [(D_u + D_v)k^2 - f_u - g_v]\lambda + D_u D_v k^4 - k^2(D_v f_u + D_u g_v) + f_u g_v - f_v g_u = 0.$$

In the absence of diffusion ($D_u = D_v = 0$), this reduces to

$$\lambda^2 - [f_u + g_v]\lambda + f_u g_v - f_v g_u = 0,$$

and the requirement that the homogeneous state is stable in the absence of diffusion leads to the conditions

$$f_u + g_v < 0, \quad f_u g_v - f_v g_u > 0. \tag{9.2.4}$$

On the other hand, the requirement that the stationary state is unstable to perturbations in the presence of diffusion means that there exists a nonzero wavenumber k for which $\lambda(k) = 0$. Setting $\lambda = 0$ in the quadratic equation yields

$$D_u D_v k^4 - k^2(D_v f_u + D_u g_v) + f_u g_v - f_v g_u = 0,$$

which will have a positive solution for k^2 provided that

$$D_v f_u + D_u g_v > 0. \tag{9.2.5}$$

Equations (9.2.4) and (9.2.5) give the conditions for a Turing instability. One can use these to identify regions in parameter space where a Turing instability can occur. Suppose that one chooses a point in parameter space just outside this instability region such that plotting $\mathrm{Re}[\lambda(k)]$ as a function of k yields dispersion curves that lie below the horizontal axis. However, if one varies an appropriate bifurcation parameter so that one crosses a boundary of the instability region, then at least one of the dispersion curves crosses the axis at a critical wavenumber k_c, and spatially periodic patterns at the critical wavelength $2\pi/k_c$ start to grow; this is the onset of the Turing instability. Finally, note that in the given bounded domain, the linear operator $\mathbb{L}$ is compact and has a discrete spectrum. In other words, the allowed wave vectors are discrete with

$$k = \frac{\pi}{L}\sqrt{n_x^2 + n_y^2}$$

for integers n_x, n_y. However, when discussing dispersion curves one often treats k as a continuous variable.

Example 9.1. In order to illustrate the above ideas, we consider a nondimensionalized RD system analyzed by Barrio et al. [24]:

$$\frac{\partial u}{\partial t} = D\nabla^2 u + \kappa(u + av - uv^2 - Cuv) \tag{9.2.6a}$$

$$\frac{\partial u}{\partial t} = \nabla^2 u + \kappa(-u + bv + uv^2 + Cuv). \tag{9.2.6b}$$

This has a unique stationary state at $(u, v) = (0, 0)$. Linearizing about this state leads to the characteristic equation

$$\left| \begin{pmatrix} \kappa - Dk^2 & \kappa a \\ -\kappa & b\kappa - k^2 \end{pmatrix} - \lambda(k)\mathbf{I} \right| = 0, \tag{9.2.7}$$

which yields the quadratic equation

$$(\kappa - Dk^2 - \lambda)(b\kappa - k^2 - \lambda) + a\kappa^2 = 0.$$

When $k = 0$, we have

$$\lambda^2 - (1+b)\kappa\lambda + (b+a)\kappa^2 = 0,$$

with roots

$$\lambda = \frac{\kappa}{2}\left[(1+b) \pm \sqrt{(1+b)^2 - 4(b+a)}\right].$$

The fixed point $(0,0)$ is stable to uniform perturbations provided $b < -1$ and $b+a > 0$. The condition for a Turing instability is that there exists a positive solution for k^2 when $\lambda(k) = 0$:

$$Dk^4 - \kappa k^2(Db+1) + \kappa^2(b+a) = 0.$$

At onset of the instability, the discriminant of the quadratic for k^2 vanishes, that is, $[\kappa(Db+1)]^2 = 4D\kappa^2(b+a)$, which implies

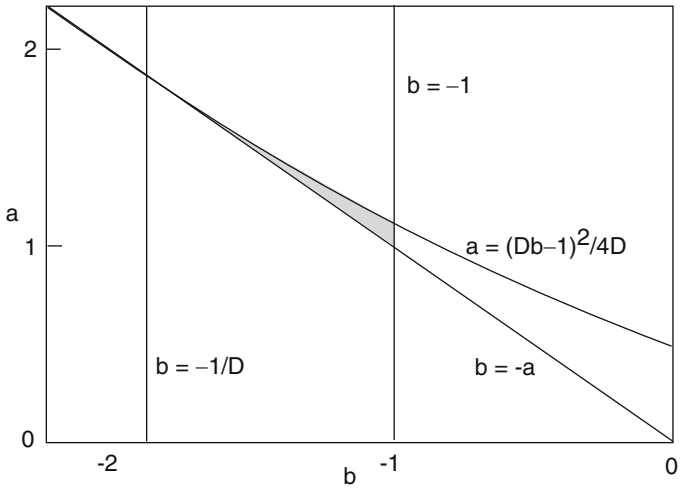


Fig. 9.6: Stability diagram for the example RD system given by Eq. (9.2.6). The *shaded region* indicates where in parameter space the homogeneous fixed point $(0,0)$ undergoes a Turing instability

$$a = \frac{(Db-1)^2}{4D},$$

and the critical wavenumber is given by

$$k_c^2 = \frac{\kappa(Db+1)}{2D}.$$

Such a solution exists provided that $Db+1>0$. Note that the parameter C plays no role in the linear analysis. The Turing instability region in the (a,b)-plane can now be determined, as illustrated in Fig. 9.6. One striking feature is that the Turing instability occupies a relatively small domain of parameter space—this is a common finding in the theory of Turing pattern formation and is known as the fine-tuning problem. One way to broaden the domain is to include the effects of noise (see Sect. 9.3). In Fig. 9.7 we sketch the dominant dispersion curves for two different sets of parameters, indicating a small band of unstable eigenmodes in both cases.

9.2.2 Amplitude Equations and Bifurcation Theory

So far we used linear theory to show how a stationary state of the RD system (9.2.1) can undergo a pattern forming instability, leading to the growth of some linear combination of eigenmodes of the underlying linear operator. However, as the eigenmodes increase in amplitude, the linear approximation breaks down and nonlinear theory is necessary in order to investigate whether or not a stable pattern ultimately forms [666]. Suppose that we take μ to denote a bifurcation parameter, such that the stationary state is stable for $\mu<\mu_c$ and undergoes a Turing instability at $\mu=\mu_c$. (In the model example given by Eq. (9.2.6), we could identify μ with the parameter a, e.g.) Sufficiently close to the bifurcation point, we can treat $\mu-\mu_c=\varepsilon$ as a small parameter and carry out a perturbation expansion in powers of ε. This generates a

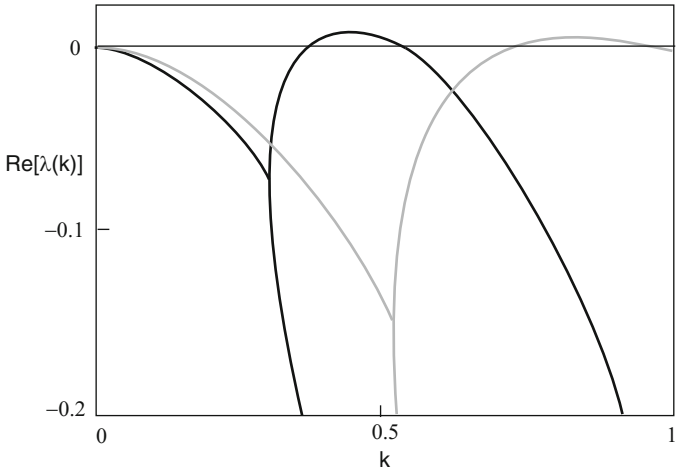


Fig. 9.7: The dispersion relation $\lambda(k)$ of the dominant eigenvalue for two different parameter sets. In both cases there is a small band of wavenumbers for which $\mathrm{Re}[\lambda(k)]>0$. The critical wavenumber k_c at the peak of the dispersion curve determines the wavelength of the emerging pattern

dynamical equation for the amplitude of the pattern that can be used to investigate pattern stability, at least in the weakly nonlinear regime. However, one immediate difficulty in carrying out this program is that a large number of eigenmodes could be excited beyond the bifurcation point. In the case of an unbounded domain, the wavenumber k is continuous-valued, since $\mathbb{L}$ now has a continuous spectrum, so that there will be a continuous band of growing modes in a neighborhood of $k = k_c$ when $\mu > \mu_c$. Moreover, the unbounded system is symmetric with respect to the action of the Euclidean group $\mathbb{E}(2)$—the group of rigid body translations, rotations, and reflections in the plane. (See Box 9A for a brief primer on group theory and its application to bifurcation theory.) This means that all eigenmodes $e^{i\mathbf{k}\cdot\mathbf{x}}$ lying on the critical circle $|\mathbf{k}| = k_c$ will be excited. Even though the number of excited eigenmodes becomes finite in a bounded domain, it can still be very large when the size of the domain satisfies $L \gg 2\pi/k_c$.

Box 9A. Symmetric bifurcation theory and pattern formation [225].

Group axioms. A group Γ is a set of elements $a \in \Gamma$ together with a group operation $\cdot$ that satisfies the following axioms: (i) If $a, b \in \Gamma$, then $a \cdot b \in \Gamma$ (closure). (ii) For all $a, b, c \in \Gamma$, we have $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ (associativity). (iii) There exists an identity element 1 such that for all $a \in \Gamma$, we have $1 \cdot a = a \cdot 1 = a$ (identity element). (iv) For each $a \in \Gamma$, there exists an element a^{-1} such that $a \cdot a^{-1} = a^{-1} \cdot a = 1$ (inverse element).

Group representations. A representation of a group Γ acting on an n-dimensional vector space V is a map $\rho : G \to GL(V)$, where $GL(V)$ is the general linear group on V, such that

$$\rho(a \cdot b) = \rho(a)\rho(b), \text{ for all } a, b \in \Gamma.$$

For a particular choice of basis set for V, $GL(V)$ can be identified with the group of invertible $n \times n$ matrices. An *irreducible representation* is one that has no proper closed sub-representations. In the language of matrices, this means that it is not possible to choose a basis set in which the matrix representation can be written in block diagonal form

$$D(a) = \begin{pmatrix} D^{(1)}(a) & 0 & \dots 0 & \\ 0 & D^{(2)}(a) \dots 0 & & \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & D^{(k)}(a) \end{pmatrix} = D^{(1)}(a) \oplus D^{(2)}(a) \dots \oplus D^{(k)}(a),$$

where $D^{(j)}(a)$ are submatrices.

Group action on a function space. Suppose that the map ρ is a representation of a group Γ acting on a finite-dimensional vector space V. Let $u : V \to \mathbb{R}$ be a function mapping elements of V to the real line. (One

could also consider complex-valued functions.) There is then a natural representation σ of the group Γ acting on the space of functions $C(V,\mathbb{R})$:

$$\sigma(\gamma)\cdot u(\mathbf{x}) = u(\rho(\gamma)^{-1}(\mathbf{x})), \text{ for all } \mathbf{x}\in V \text{ and } \gamma\in\Gamma.$$

It is necessary to use the inverse element $\rho(\gamma)^{-1}$ to ensure that σ is a group representation. That is,

$$\begin{aligned}\sigma(\gamma_1)\cdot\sigma(\gamma_2)\cdot u(\mathbf{x}) &= \sigma(\gamma_1)\cdot u(\rho(\gamma_2)^{-1}(\mathbf{x})) = \sigma(\gamma_1)\cdot u_2(\mathbf{x})\\ &= u_2(\rho(\gamma_1)^{-1}(\mathbf{x})) = u(\rho(\gamma_2)^{-1}\rho(\gamma_1)^{-1}(\mathbf{x}))\\ &= u([\rho(\gamma_1)\rho(\gamma_2)]^{-1}(\mathbf{x})) = u([\rho(\gamma_1\gamma_2)]^{-1}(\mathbf{x}))\\ &= \sigma(\gamma_1\gamma_2)\cdot u(\mathbf{x}).\end{aligned}$$

In the following we will use the same symbol γ for an abstract group element and its corresponding group representation.

Equivariance. Suppose that $u\in C(\mathbb{R}^2,\mathbb{R})$ is the solution to a scalar reaction-diffusion equation of the form

$$\frac{\partial u(\mathbf{x},t)}{\partial t} = \nabla^2_{\mathbf{x}} u(\mathbf{x},t) + f(u(\mathbf{x},t)).$$

(One could equally well consider a system of RD equations.) This equation is equivariant with respect to the natural action of the Euclidean group $\mathbb{E}(2)$ on $C(\mathbb{R}^2,\mathbb{R})$. That is, if $u(\mathbf{x},t)$ is a solution, then so is $u(\gamma^{-1}\mathbf{x},t)$ for all $\gamma\in\mathbb{E}(2)$. This is a consequence of the fact that the Laplacian operator is invariant with respect to the action of $\mathbb{E}(2)$: $\nabla^2_{\gamma^{-1}\mathbf{x}} = \nabla^2_{\mathbf{x}}$. In other words, the operators γ and ∇^2 commute, $\gamma\nabla^2 = \nabla^2\gamma$. Equivariance then follows, since

$$\begin{aligned}0 &= \gamma\left[\frac{\partial u(\mathbf{x},t)}{\partial t} - \nabla^2_{\mathbf{x}} u(\mathbf{x},t) - f(u(\mathbf{x},t))\right]\\ &= \frac{\partial u(\gamma^{-1}\mathbf{x},t)}{\partial t} - \nabla^2_{\gamma^{-1}\mathbf{x}} u(\gamma^{-1}\mathbf{x},t) - f(u(\gamma^{-1}\mathbf{x},t))\\ &= \frac{\partial u(\gamma^{-1}\mathbf{x},t)}{\partial t} - \nabla^2_{\mathbf{x}} u(\gamma^{-1}\mathbf{x},t) - f(u(\gamma^{-1}\mathbf{x},t)).\end{aligned}$$

Equivariance has major implications for the bifurcation structure of solutions to the PDE. In particular, suppose that u_0 is a homogeneous stationary solution. Such a solution preserves full Euclidean symmetry so that the PDE obtained by linearizing about the stationary solution is also equivariant with respect to $\mathbb{E}(2)$. Writing the linear PDE as $\partial_t u = \mathbb{L}u$, where $\mathbb{L}$ is a linear operator on $C(\mathbb{R}^2,\mathbb{R})$, we see that $\mathbb{L}$ commutes with the group elements γ. This implies that generically the eigenfunctions of

$\mathbb{L}$ form irreducible representations of $\mathbb{E}(2)$. (Strictly speaking $\mathbb{L}$ has a continuous rather than a discrete spectrum. One could restrict the PDE to a bounded domain to obtain a discrete spectrum, but it would have to be sufficiently large so that Euclidean symmetry still approximately holds.) Applying the group action to the eigenvalue equation for $\mathbb{L}$ shows that

$$\begin{aligned} 0 &= \gamma \cdot [\mathbb{L}\phi_\lambda(\mathbf{x}) - \lambda\phi_\lambda(\mathbf{x})] \\ &= \mathbb{L}\gamma\phi_\lambda(\mathbf{x}) - \lambda\gamma\phi_\lambda(\mathbf{x}) \\ &= \mathbb{L}\phi_\lambda(\gamma^{-1}\mathbf{x}) - \lambda\phi_\lambda(\gamma^{-1}\mathbf{x}). \end{aligned}$$

Thus $\phi_\lambda(\gamma^{-1}\mathbf{x})$ for all $\gamma \in \Gamma$ have the same eigenvalue λ. This degeneracy is an immediate consequence of the underlying symmetry. In general, we don't expect any further degeneracy for the given eigenvalue, so the eigenfunctions $\phi_\lambda(\gamma^{-1}\mathbf{x})$ form an irreducible representation of the group.

The same analysis holds if we restrict solutions to doubly periodic functions, except that the resulting PDE is now equivariant with respect to the discrete group $\Gamma = D_n \dot{+} \mathbf{T}^2$. As we show below, bifurcations from a homogeneous stationary solution can now be analyzed in terms of a system of ODEs—amplitude equations—of the form

$$\dot{\mathbf{z}} = F(\mathbf{z}) \tag{9.2.8}$$

where $\mathbf{z}, F(\mathbf{z}) \in V$ with $V = \mathbb{R}^n$ or $\mathbb{C}^n$. These equations are also equivariant with respect to Γ, since

$$\gamma \cdot F(z) = F(\gamma \cdot z)$$

for all $\gamma \in \Gamma$. It immediately follows that if $z(t)$ is a solution to the system of ODEs, then so is $\gamma \cdot z(t)$. Moreover, since $F(0) = 0$, the origin is an equilibrium that is invariant under the action of the full symmetry group Γ. Thus linearizing about the fixed point $\mathbf{z} = 0$ generates a linear operator whose eigenvectors form irreducible representations of the group Γ.

Isotropy subgroups. The symmetries of any particular equilibrium solution $\mathbf{z}$ form a subgroup called the *isotropy* subgroup of $\mathbf{z}$ defined by

$$\Sigma_{\mathbf{z}} = \{\sigma \in \Gamma : \ \sigma\mathbf{z} = \mathbf{z}\}. \tag{9.2.9}$$

More generally, we say that Σ is an isotropy subgroup of Γ if $\Sigma = \Sigma_{\mathbf{z}}$ for some $\mathbf{z} \in V$. Isotropy subgroups are defined up to some conjugacy. A group Σ is conjugate to a group $\hat{\Sigma}$ if there exists $\sigma \in \Gamma$ such that $\hat{\Sigma} = \sigma^{-1}\Sigma\sigma$. The *fixed point subspace* of an isotropy subgroup Σ, denoted by $\mathrm{Fix}(\Sigma)$, is the set of points $\mathbf{z} \in V$ that are invariant under the action of Σ,

$$\mathrm{Fix}(\Sigma) = \{\mathbf{z} \in V : \ \sigma\mathbf{z} = \mathbf{z} \ \forall \ \sigma \in \Sigma\}. \tag{9.2.10}$$

Finally, the *group orbit* through a point $\mathbf{z}$ is

$$\Gamma\mathbf{z} = \{\sigma\mathbf{z} : \ \sigma \in \Gamma\}. \tag{9.2.11}$$

If $\mathbf{z}$ is an equilibrium solution of equation (9.2.8), then so are all other points of the group orbit (by equivariance). One can now adopt a strategy that restricts the search for solutions of equation (9.2.8) to those that are fixed points of a particular isotropy subgroup. In general, if a dynamical system is equivariant under some symmetry group Γ and has a solution that is a fixed point of the full symmetry group, then we expect a loss of stability to occur upon variation of one or more system parameters. Typically such a loss of stability will be associated with the occurrence of new solution branches with isotropy subgroups Σ smaller than Γ. One says that the solution has spontaneously broken symmetry from Γ to Σ. Instead of a unique solution with the full set of symmetries Γ a set of symmetrically related solutions (orbits under Γ modulo Σ) each with symmetry group (conjugate to) Σ is observed.

Equivariant branching lemma. Suppose that the system of equations (9.2.8) has a fixed point of the full symmetry group Γ. The *equivariant branching lemma* [225] states that generically there exists a (unique) equilibrium solution bifurcating from the fixed point for each of the axial subgroups of Γ under the given group action—a subgroup $\Sigma \subset \Gamma$ is *axial* if $\dim \mathrm{Fix}(\Sigma) = 1$. The heuristic idea underlying this lemma is as follows. Let Σ be an axial subgroup and $\mathbf{z} \in \mathrm{Fix}(\Sigma)$. Equivariance of F then implies that

$$\sigma F(\mathbf{z}) = F(\sigma\mathbf{z}) = F(\mathbf{z}) \tag{9.2.12}$$

for all $\sigma \in \Sigma$. Thus $F(\mathbf{z}) \in \mathrm{Fix}(\Sigma)$ and the system of coupled ODEs (9.2.8) can be reduced to a single equation in the fixed point space of Σ. Such an equation is expected to support a codimension one bifurcation, in which new stationary solutions emerge whose amplitudes correspond to fixed points of axial isotropy subgroups. Since the codimension of a bifurcation corresponds generically to the number of parameters that need to be varied in order to induce the bifurcation, one expects the primary bifurcations to be codimension one. Thus one can systematically identify the various expected primary bifurcation branches by constructing the associated axial subgroups and finding their fixed points.

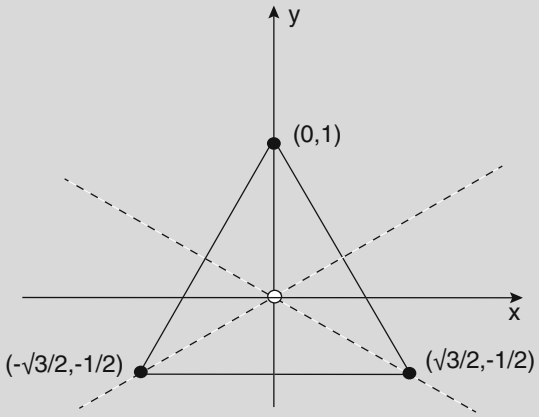


Fig. 9.8: Symmetry operations on an equilateral triangle

Example. For the sake of illustration, consider the full symmetry group D_3 of an equilateral triangle (see Fig. 9.8). The action is generated by the matrices (in an appropriately chosen orthonormal basis)

$$R = \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix}, \quad S = \begin{pmatrix} -1 & 0 \\ 0 & 1. \end{pmatrix}$$

Here R is a rotation by $\pi/3$ and S is a reflection about the y-axis. (The generators of a discrete group form a minimal set of group elements from which all other group elements can be obtained by combinations of group operations.) Clearly, R fixes only the origin, while S fixes any point $(x,0)$. We deduce that the isotropy subgroups are as follows: (i) the full symmetry group D_3 with single fixed point $(0,0)$; (ii) the two-element group $\mathbf{Z}_2(S)$ generated by S, which fixes the x-axis, and the groups that are conjugate to $\mathbf{Z}_2(S)$ by the rotations R and R^2; and (iii) a trivial group formed by the identity matrix in which every point is a fixed point. The isotropy subgroups form the hierarchy

$$\{I\} \subset \mathbf{Z}_2(S) \subset D_3.$$

It follows that up to conjugacy the only axial subgroup is $\mathbf{Z}_2(S)$. Thus we expect the fixed point $(0,0)$ to undergo a symmetry-breaking bifurcation to an equilibrium that has reflection symmetry. Such an equilibrium will be given by one of the three points $\{(x,0), R(x,0), R^2(x,0)\}$ on the group orbit generated by discrete rotations. Which of these states is selected will depend on initial conditions, that is, the broken rotation symmetry is hidden. Note that a similar analysis can be carried out for the symmetry group D_4 of the square (see Ex. 9.2). Now, however, there are two distinct

types of reflection axes: those joining the middle of opposite edges and those joining opposite vertices. Since these two types of reflections are not conjugate to each other, there are now two distinct axial subgroups.

One common property shared by many biological patterns observed in nature is that the primary bifurcations tend to generate relatively simple patterns such as stripes and hexagonal spot patterns. This motivates the mathematical simplification of restricting the space of solutions of the RD system (9.2.1) to that of *doubly periodic functions* . That is, one imposes the conditions

$$u(\mathbf{x}+\ell,t) = u(\mathbf{x},t), \quad v(\mathbf{x}+\ell,t) = v(\mathbf{x},t)$$

for every $\ell \in \mathscr{L}$ where $\mathscr{L}$ is some regular planar lattice. The lattice $\mathscr{L}$ is generated by two linearly independent vectors ℓ_1 and ℓ_2:

$$\mathscr{L} = \{(m_1\ell_1 + m_2\ell_2) : \ m_1, m_2 \in \mathbf{Z}\} \tag{9.2.13}$$

with lattice spacing $d = |\ell_j|$. Let ψ be the angle between the two basis vectors ℓ_1 and ℓ_2. We can then distinguish three types of lattice according to the value of ψ: square lattice ($\psi = \pi/2$), rhombic lattice ($0 < \psi < \pi/2$, $\psi \neq \pi/3$), and hexagonal ($\psi = \pi/3$) (see Table 9.1). Restriction to double periodicity means that the original Euclidean symmetry group is now restricted to the symmetry group of the lattice, $\Gamma = D_n \dot{+} \mathbf{T}^2$, where D_n is the *holohedry* of the lattice, the subgroup of rotations and reflections $\mathbf{O}(2)$ that preserves the lattice, and $\mathbf{T}^2$ is the two torus of planar translations modulo the lattice. Thus, the holohedry of the rhombic lattice is D_2, the holohedry of the square lattice is D_4, and the holohedry of the hexagonal lattice is D_6, (see Fig. 9.9). There are only a finite number of rotations and reflections to consider for each lattice (modulo an arbitrary rotation of the whole plane). Consequently, there is only a finite set of candidate excited eigenmodes.

Imposing double periodicity on the marginally stable eigenmodes restricts the lattice spacing such that the critical wavevector $\mathbf{k}_c$ lies on the *dual lattice* $\widehat{\mathscr{L}}$; the generators of the dual lattice satisfy $\hat{\ell}_i.\ell_j = \delta_{i,j}$ for $i,j = 1,2$. In order to generate the simplest observed patterns, d is chosen so that k_c is the shortest length of a dual wave vector. Linear combinations of eigenmodes that generate doubly periodic solutions corresponding to dual wave vectors of shortest length are then given by $\mathbf{c}(\mathbf{x}) = \mathbf{c}_0\phi(\mathbf{x})$ where

$$\phi(\mathbf{x}) = \sum_{j=1}^{N} z_j e^{i\mathbf{k}_j\cdot\mathbf{x}} + c.c \tag{9.2.14}$$

where the z_j are complex amplitudes. Here $N = 2$ for the square lattice with $\mathbf{k}_1 = \mathbf{k}_c$ and $\mathbf{k}_2 = R_{\pi/2}\mathbf{k}_c$, where R_ξ denotes rotation through an angle ξ. Similarly, $N = 3$ for the hexagonal lattice with $\mathbf{k}_1 = \mathbf{k}_c$, $\mathbf{k}_2 = R_{2\pi/3}\mathbf{k}_c$, and $\mathbf{k}_3 = R_{4\pi/3}\mathbf{k}_c = -\mathbf{k}_1 - \mathbf{k}_2$. It follows that the space of marginally stable eigenmodes can be identified with the N-dimensional complex vector space spanned by the vectors $(z_1,\ldots,z_N) \in \mathbb{C}^N$ with

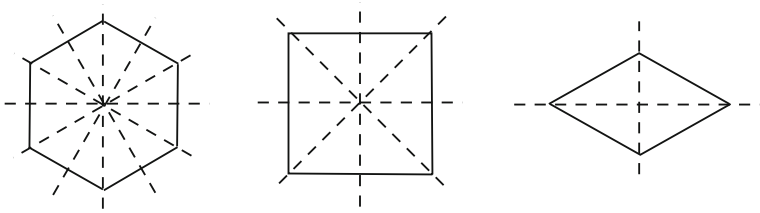


Fig. 9.9: Holohedries of the plane

$N = 2$ for square or rhombic lattices and $N = 3$ for hexagonal lattices. It can be shown that these form irreducible representations of the group $\Gamma = D_n \dot{+} \mathbf{T}^2$ (see Box 9A for a definition of irreducibility), whose action on $\mathbb{C}^N$ is induced by the corresponding action of Γ on $\phi(\mathbf{x})$. For example, on a hexagonal lattice, a translation $\phi(\mathbf{x}) \to \phi(\mathbf{r} - \mathbf{s})$ induces the action

$$\gamma \cdot (z_1, z_2, z_3) = (z_1 \mathrm{e}^{-i\theta_1}, z_2 \mathrm{e}^{-i\theta_2}, z_3 \mathrm{e}^{i(\theta_1+\theta_2)}) \tag{9.2.15}$$

with $\theta_j = \mathbf{k}_j \cdot \mathbf{s}$, a rotation $\phi(\mathbf{x}) \to \phi(R_{-2\pi/3}\mathbf{x})$ induces the action

$$\gamma \cdot (z_1, z_2, z_3) = (z_3, z_1, z_2), \tag{9.2.16}$$

Lattice	ℓ_1	ℓ_2	$\hat{\ell}_1$	$\hat{\ell}_2$
Square	$(1,0)$	$(0,1)$	$(1,0)$	$(0,1)$
Hexagonal	$(1,0)$	$\frac{1}{2}(1,\sqrt{3})$	$(1,\frac{-1}{\sqrt{3}})$	$(0,\frac{2}{\sqrt{3}})$
Rhombic	$(1,0)$	$(\cos\eta, \sin\eta)$	$(1,-\cot\eta)$	$(0,\csc\eta)$

Table 9.1: Generators for the planar lattices and their dual lattices in the case of unit lattice spacing ($d = 1$)

and a reflection across the x-axis [assuming $\mathbf{k}_c = k_c(1,0)$] induces the action

$$\gamma \cdot (z_1, z_2, z_3) = (z_1, z_3, z_2). \tag{9.2.17}$$

The full action of $D_n \dot{+} \mathbf{T}^2$ on $\mathbb{C}^N$ for the various regular planar lattices is given in Table 9.2.

It turns out that symmetry plays a major role in the selection and stability of patterns that emerge via a Turing instability (see Box 9A). Here we will describe the role of symmetry from the perspective of amplitude equations. The latter refer to the system of nonlinear ODEs describing the slow dynamics of the amplitudes $\mathbf{z} = (z_1, \ldots, z_N)$ close to the bifurcation point, which can be analyzed using the method of multiple scales. The basic idea is that close to the Turing bifurcation point where $\mu - \mu_c = \varepsilon$, the critical eigenvalue $\lambda(k_c) = O(\varepsilon)$, which means that the excited

eigenmodes grow slowly with respect to time. (We are assuming that $\lambda(k_c)$ is real; if it has a nonzero imaginary part, then a Turing–Hopf instability may lead to the formation of oscillatory patterns; see also Box 4B.) In order to pick out this slow exponential growth using perturbation theory, one introduces a slow time variable $\tau = \varepsilon t$ and substitutes the series expansion

$$\mathbf{u} = \mathbf{u}^* + \varepsilon^{1/2}\mathbf{u}_1(\mathbf{x},\tau) + \varepsilon^{1/2}\mathbf{u}_2(\mathbf{x},\tau) + \varepsilon^{3/2}\mathbf{u}_3(\mathbf{x},\tau) + \ldots$$

into the full RD system (9.2.1). Taylor expanding the nonlinear functions $f(u,v)$ and $g(u,v)$ about the stationary solution (u^*,v^*) and collecting terms having the same power of ε leads to a hierarchy of equations of the general form

$$\mathbb{L}u_n(\mathbf{x},\tau) = h_n(u_1,\ldots u_{n-1}),$$

D_2	Action	D_4	Action	D_6	Action
$\mathbf{1}$	(z_1,z_2)	$\mathbf{1}$	(z_1,z_2)	$\mathbf{1}$	(z_1,z_2,z_3)
ξ	(z_1^*,z_2^*)	ξ	(z_2^*,z_1)	ξ	(z_2^*,z_3^*,z_1^*)
κ_η	(z_2,z_1)	ξ^2	(z_1^*,z_2^*)	ξ^2	(z_3,z_1,z_2)
$\kappa_\eta\xi$	(z_2^*,z_1^*)	ξ^3	(z_2,z_1^*)	ξ^3	(z_1^*,z_2^*,z_3^*)
		κ	(z_1,z_2^*)	ξ^4	(z_2,z_3,z_1)
		$\kappa\xi$	(z_2^*,z_1^*)	ξ^5	(z_3^*,z_1^*,z_2^*)
		$\kappa\xi^2$	(z_1^*,z_2)	κ	(z_1,z_3,z_2)
		$\kappa\xi^3$	(z_2,z_1)	$\kappa\xi$	(z_2^*,z_1^*,z_3^*)
				$\kappa\xi^2$	(z_3,z_2,z_1)
				$\kappa\xi^3$	(z_1^*,z_3^*,z_2^*)
				$\kappa\xi^4$	(z_2,z_1,z_3)
				$\kappa\xi^5$	(z_3^*,z_2^*,z_1^*)

Table 9.2: (*Left*) $D_2{+}\mathbf{T}^2$ action on rhombic lattice; (*center*) $D_4{+}\mathbf{T}^2$ action on square lattice; (*right*) $D_6{+}\mathbf{T}^2$ action on hexagonal lattice. In each case the generators of D_n are a reflection and a rotation. For the square and hexagonal lattices, the generator κ represents reflection across the x-axis, whereas for the rhombic lattice, the generator κ_η represents reflections across the major diagonal. The CCW rotation generator ξ represents rotation through the angles π (rhombic), $\frac{\pi}{2}$ (square), and $\frac{\pi}{3}$ (hexagonal)

where $\mathbb{L}$ is the linear operator defined in Eq. (9.2.2) and h_n is a function of lower-order terms in the hierarchy. Since $h_1 \equiv 0$, it follows that the $O(\varepsilon^{1/2})$ solution u_1 is given by Eq. (9.2.14) with time-dependent amplitudes:

$$\mathbf{u}_1(\mathbf{x},\tau) = \mathbf{c}_0 \sum_{j=1}^{N} z_j(\tau) e^{i\mathbf{k}_j\cdot\mathbf{x}} + c.c.$$

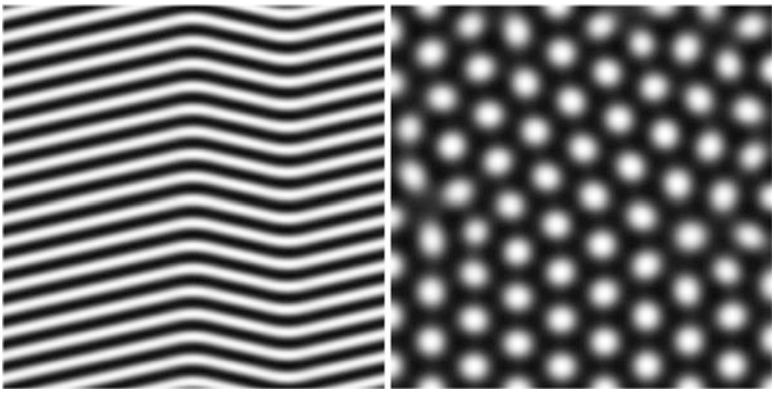


Fig. 9.10: Illustration of a *stripe pattern* and a hexagonal *spot pattern* in the plane

Applying the Fredholm alternative theorem (Box 9B) to the inhomogeneous higher-order equations then determines an amplitude equation for $\mathbf{z}(\tau)$ of the form

$$\frac{dz_j}{d\tau} = F_j(\mathbf{z}), \quad j = 1,\ldots,N, \tag{9.2.18}$$

where F_j can be expanded as a polynomial in the z_j's—close to the bifurcation point it is often sufficient to truncate the polynomials at cubic order in the amplitudes. In Ex. 9.3 we develop the steps necessary to derive the amplitude equation in the case of a simpler problem, namely, the RD system (9.2.6) defined on a ring—in this case the underlying symmetry group is $\mathbf{O}(2)$, consisting of rotations and reflections on the circle (see also Ex. 9.4). Note that although the parameter C does not appear in the linear theory, it does appear in the amplitude equation and thus, in the 2D case, determines whether or not the emerging pattern is a stripe pattern or a hexagonal spot pattern [24] (see Fig. 9.10).

Box 9B. Fredholm alternative theorem (linear operators).

The Fredholm alternative theorem for matrices (see Box 7B) can be extended to the case of linear operators acting on infinite-dimensional function spaces [554]. Consider the set of real functions $f(x)$ for $x \in [a,b]$. This is a vector space over the set of real numbers: given two functions $f_1(x), f_2(x)$ and two real numbers a_1, a_2, we can form the sum $f(x) = a_1 f_1(x) + a_2 f_2(x)$ such that $f(x)$ is also a function on $\mathbb{R}$. Suppose that we restrict the space of functions to those that are normalizable: $\int_a^b |f(x)|^2 dx < \infty$. More precisely, we are considering the Hilbert space of functions $L^2([a,b])$. We define the inner product of any two normalizable functions f, g according to

$$\langle f, g\rangle = \int_a^b f(x)g(x)dx.$$

Suppose that $\mathscr{L}$ is a linear differential operator acting on a subspace of $L^2([a,b])$ consisting of functions that are differentiable to the appropriate order, which we denote by the domain $\mathscr{D}(\mathscr{L})$. Linearity of the operator means that for $f_1, f_2 \in \mathscr{D}(\mathscr{L})$ and $a_1, a_2 \in \mathbb{R}$,

$$\mathscr{L}(a_1 f_1 + a_2 f_2) = a_1 \mathscr{L} f_1 + a_2 \mathscr{L} f_2.$$

Given the standard inner product on $L^2(\mathbb{R})$, we define the adjoint linear operator $\mathscr{L}^\dagger$ according to

$$\langle f, \mathscr{L} g\rangle = \langle \mathscr{L}^\dagger f, g\rangle, \quad f, g \in \mathscr{D}(L).$$

The operator is said to be self-adjoint if $\mathscr{L}^\dagger = \mathscr{L}$. Note that, in practice, one determines $\mathscr{L}^\dagger$ using integration by parts. For functions defined on finite intervals, this generates boundary terms that only vanish if appropriate boundary conditions are imposed. In general, this can result in different domains for $\mathscr{L}$ and $\mathscr{L}^\dagger$. Therefore, the condition for self-adjointness becomes $\mathscr{L} = \mathscr{L}^\dagger$ and $\mathscr{D}(\mathscr{L}) = \mathscr{D}(\mathscr{L}^\dagger)$. Given a differential operator $\mathscr{L}$ on $L^2([a,b])$, we can now state an infinite-dimensional version of the Fredholm alternative theorem. The inhomogeneous equation

$$\mathscr{L} f = h$$

has a solution if and only if

$$\langle h, u\rangle = 0 \quad \text{for all u satisfying} \quad \mathscr{L}^\dagger u = 0.$$

This version is useful when deriving amplitude equations for reaction-diffusion models.

Although perturbation theory is needed to determine the specific model-dependent coefficients of the cubic, the general structure of the amplitude equation can be determined from its equivariance under the action of the symmetry group $\Gamma = D_n \dot{+} \mathbf{T}^2$. In particular, it can be shown that the cubic amplitude equations for the three types of planar lattice are of the following general form [225]:

Square or rhombic lattice. First, consider planforms corresponding to a bimodal structure of the square or rhombic type ($N = 2$). Take $\mathbf{k}_1 = k_c(1,0)$ and $\mathbf{k}_2 = k_c(\cos(\varphi), \sin(\varphi))$, with $\varphi = \pi/2$ for the square lattice and $0 < \varphi < \pi/2$, $\varphi \neq \pi/3$ for a rhombic lattice. The amplitudes evolve according to a pair of equations of the form

$$\frac{dz_1}{d\tau} = z_1\left[1-\gamma_0|z_1|^2-\gamma(\varphi)|z_2|^2\right] \tag{9.2.19a}$$

$$\frac{dc_2}{d\hat{\tau}} = z_2\left[1-\gamma_0|z_2|^2-\gamma(\varphi)|z_1|^2\right] \tag{9.2.19b}$$

with $\gamma(\pi/2)=\gamma_0$. If $\gamma(\varphi)>0$, then three types of steady states are possible.

1. The homogeneous state: $z_1=z_2=0$.
2. Rolls or stripe patterns: $z_1=\sqrt{1/\gamma_0}e^{i\psi_1}, z_2=0$ or $z_1=0, z_2=\sqrt{1/\gamma_0}e^{i\psi_2}$.
3. Squares or rhombics: $z_j=\sqrt{1/[\gamma_0+\gamma(\varphi)]}e^{i\psi_j}$, $j=1,2$,

where ψ_1 and ψ_2 are arbitrary phases. A standard linear stability analysis shows that if $\gamma(\varphi)>\gamma_0$, then rolls are stable whereas the square or rhombic patterns are unstable. The opposite holds if $\gamma(\varphi)<\gamma_0$. Note that here stability is defined with respect to perturbations with the same lattice structure.

Hexagonal lattice. Next consider planforms on a hexagonal lattice with $N=3$, $\varphi_1=0$, $\varphi_2=2\pi/3$, and $\varphi_3=-2\pi/3$. The cubic amplitude equations take the form

$$\frac{dz_j}{d\tau} = z_j\left[1-\gamma|z_j|^2-\alpha(|z_{j+1}|^2+|z_{j-1}|^2)\right]+\beta z_{j+1}^* z_{j-1}^* \tag{9.2.20}$$

where $j=1,2,3$ mod 3. Unfortunately, Eq. (9.2.20) is not sufficient to determine the selection and stability of the steady-state solutions bifurcating from the homogeneous state. One has to carry out an *unfolding* of the amplitude equation that includes higher-order terms (quartic and quintic) in $z,\bar{z}$ whose general form can also be deduced using symmetry arguments.

9.3 Stochastic Pattern Formation and the RD Master Equation

As in the case of non-spatial gene networks (Chap. 6) and spatial concentration gradients (Sect. 9.1), one needs to take into account the possible effects of intrinsic noise due to low copy numbers on spontaneous pattern formation. There have been a number of recent studies that have incorporated diffusion into a stochastic biochemical network by discretizing space and treating diffusion as a hopping reaction, which can then be represented in terms of a generalized reaction–diffusion master equation [47, 87, 88, 399, 429, 573, 689]. One of the interesting implications of such studies is that intrinsic noise can increase the parameter region in which patterns form, mitigating the aforementioned problem of parameter fine-tuning. Moreover, spectral methods have been used to detect a range of noise-induced dynamical phenomena, based on a linear noise approximation of the reaction–diffusion master equation. In this section, we will develop the basic theory of stochastic pattern formation by considering the well-known example of the chemical Brusselator—applications to cell polarization will be developed in Sect. 9.5.

9.3.1 Turing Instability in the Brusselator Model

Consider a spatially extended version of the Brusselator model introduced in Sect. 6.4.5, which is given by the RD system

$$\frac{du}{dt} = D_u \nabla^2 u + a - (b+1)u + u^2 v \tag{9.3.1a}$$

$$\frac{dv}{dt} = D_v \nabla^2 v + bu - u^2 v, \tag{9.3.1b}$$

where D_u and D_v are the diffusion coefficients for u and v, respectively. Linearizing about the spatially homogeneous steady-state solution $u^* = a, v^* = b/a$ by setting

$$u(\mathbf{x},t) = u^* + \xi(\mathbf{x})\mathrm{e}^{\lambda t}, \quad v(\mathbf{x},t) = v^* + \eta(\mathbf{x})\mathrm{e}^{\lambda t}$$

and expanding to first order in ξ, η leads to the eigenvalue equation

$$\lambda \begin{pmatrix} \xi(\mathbf{x}) \\ \eta(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} b-1+D_u\nabla^2 & a^2 \\ -b & -a^2+D_v\nabla^2 \end{pmatrix} \begin{pmatrix} \xi(\mathbf{x}) \\ \eta(\mathbf{x}) \end{pmatrix}. \tag{9.3.2}$$

This has eigensolutions of the form

$$\xi(\mathbf{x}) = \hat{\xi}(k)\mathrm{e}^{-i\mathbf{k}\cdot\mathbf{x}}, \quad \eta(\mathbf{x}) = \hat{\eta}(k)\mathrm{e}^{-i\mathbf{k}\cdot\mathbf{x}},$$

such that $\nabla^2 \to -k^2$ and λ is the solution to

$$(b-1-D_u k^2 - \lambda)(a^2 + D_v k^2 + \lambda) - a^2 b = 0.$$

That is, $\lambda = \lambda_\pm(k)$ with

$$\lambda_\pm(k) = \frac{1}{2}\left[\Gamma(k) \pm \sqrt{\Gamma(k)^2 + 4\Lambda(k)}\right], \tag{9.3.3}$$

where

$$\Gamma(k) = b - 1 - a^2 - (D_u + D_v)k^2, \quad \Lambda(k) = (b-1-D_u k^2)(a^2 + D_v k^2) - a^2 b. \tag{9.3.4}$$

One of the conditions for a Turing instability is that the homogeneous fixed point should be stable with respect to homogeneous ($k = 0$) perturbations (see Sect. 9.2). Setting $k = 0$ in Eq. (9.3.3), we thus require $\mathrm{Re}[\lambda_\pm(0)] < 0$, where

$$\lambda_\pm(0) = \frac{1}{2}\left[b - 1 - a^2 \pm ia\right]. \tag{9.3.5}$$

This yields the necessary condition $b < 1 + a^2$. Now observe that as k increases from zero $\Gamma(k)$ becomes more negative. Therefore, in order that the fixed point becomes

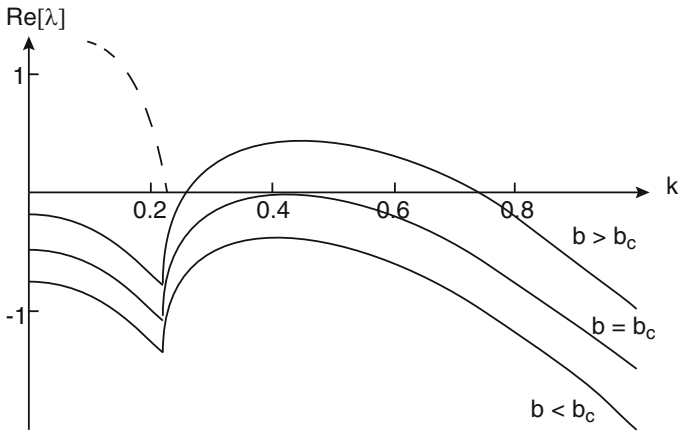


Fig. 9.11: Sketch of dispersion curves for the deterministic Brusselator model, showing $\mathrm{Re}[\lambda(k)]$ as a function of k for the most unstable eigenvalue. Parameter values are $a = 1.5, D_u = 2.8,$ *and* $D_v = 22.4$. Three different values for b are used: $b = 1.8 < b_c$, $b = b_c = 2.34$, and $b = 2.9 > b_c$. *Dashed line* shows imaginary part of $\lambda(k)$, which is approximately the same for all given b values. Redrawn from [429]

unstable due to the growth of some nonzero frequency mode ($k \neq 0$), we require $\Lambda(k)$ to become positive. The critical wavenumber k_c for a Turing instability is thus given by the condition $\Lambda(k_c) = 0$, that is,

$$[(b-1)D_v - a^2 D_u]k_c^2 = a^2.$$

Hence, a Turing instability will occur provided that

$$1 + \frac{D_u}{D_v}a^2 < b < 1 + a^2,$$

which immediately implies that $D_u < D_v$, that is, Y molecules diffuse more quickly than X molecules. If b is taken to be a bifurcation parameter, then increasing b from zero will lead to a Turing instability at the critical value $b_c = 1 + D_u a^2/D_v$, beyond which spatially periodic eigenmodes with spatial frequencies around k_c will start to grow (see Fig. 9.11). As the amplitude of these eigenmodes increases beyond the linear regime, saturating nonlinearities of the full system will typically stabilize the resulting patterns, whose fundamental wavelength will be approximately given by $2\pi/k_c$.

9.3.2 Linear-Noise Approximation of RD Master Equation

In order to incorporate diffusion into the corresponding master equation (6.4.28), we follow McKane et al. [429] by partitioning the cell into small domains with centers

at discrete lattice points ℓ. For simplicity, it is assumed that the domain centers are distributed on a regular d-dimensional lattice $\mathscr{L}$ ($d = 1,2,3$) with lattice spacing $\Delta\ell$. The spatially discrete version of the RD system (9.3.1) is

$$\frac{\partial u_\ell}{dt} = a - (b+1)u_\ell + u_\ell^2 v_\ell + \alpha_u \Delta u_\ell, \tag{9.3.6a}$$

$$\frac{\partial v_\ell}{dt} = bu_\ell - u_\ell^2 v_\ell + \alpha_v \Delta v_\ell. \tag{9.3.6b}$$

Here Δ is the discrete Laplacian defined as (see Ex. 9.5)

$$\Delta u_\ell = \frac{2}{z} \sum_{\ell' \in \partial\ell} [u_{\ell'} - u_\ell], \tag{9.3.7}$$

where $\ell' \in \partial\ell$ indicates that we are summing over nearest neighbors of ℓ on the lattice and z is the number of nearest neighbors (lattice coordination number). The continuum RD Eq. (9.5.3) is recovered in the continuum limit $\Delta l \to 0$ and $\alpha_u, \alpha_v \to \infty$ with $2\alpha_u \Delta l^2 = D_u$ and $2\alpha_v \Delta l^2 = D_v$ fixed and $u_\ell(t), v_\ell(t) \to u(\mathbf{x},t), v(\mathbf{x},t)$. Now let $n_{\ell,j}$ denote the number of X molecules ($j = 1$) and Y molecules ($j = 2$) in the domain with center at ℓ and introduce the local densities $u_{\ell,j} = n_{\ell,j}/\Omega$. A major advantage of discretizing space is that diffusion can now be represented by a set of hopping reactions and treated on an equal footing with the chemical reactions in Eq. (6.4.25). The full set of reactions is

$$(n_{\ell,1}, n_{\ell,2}) \xrightarrow{T_{\ell,1}} (n_{\ell,1}+1, n_{\ell,2}), \tag{9.3.8a}$$

$$(n_{\ell,1}, n_{\ell,2}) \xrightarrow{T_{\ell,2}} (n_{\ell,1}-1, n_{\ell,2}+1) \tag{9.3.8b}$$

$$(n_{\ell,1}, n_{\ell,2}) \xrightarrow{T_{\ell,3}} (n_{\ell,1}+1, n_{\ell,2}-1) \tag{9.3.8c}$$

$$(n_{\ell,1}, n_{\ell,2}) \xrightarrow{T_{\ell,4}} (n_{\ell,1}-1, n_{\ell,2}), \tag{9.3.8d}$$

$$(n_{\ell,j}, n_{\ell',j}) \xrightarrow{T^j_{\ell,\ell'}} (n_{\ell,j}-1, n_{\ell',j}+1), \quad \ell' \in \partial\ell, \tag{9.3.8e}$$

with

$$T_{\ell,r} = T_r(n_{\ell,1}, n_{\ell,2}) \text{ for } r = 1,2,3,4, \quad T^j_{\ell,\ell'} = \frac{2\alpha_j}{z} n_{\ell,j}, \tag{9.3.9}$$

and $\alpha_1 = \alpha_u, \alpha_2 = \alpha_v$. If N denotes the total number of lattice points, then there are $2N$ "chemical species" labeled by (ℓ, j). Moreover, we have $2zN$ hopping reactions and pN local biochemical reactions between the X and Y molecules, so that the total number of reactions is $R = (2z+p)N$. For the general two-species RD equation

$$\frac{\partial u_{\ell,i}}{\partial t} = \sum_{r=1}^{p} S_{ir} f_r(u_{\ell,1}, u_{\ell,2}) + \alpha_i \Delta u_{\ell,i}, \quad i = 1,2 \tag{9.3.10}$$

the corresponding RD master equation has the general form

$$\frac{dP(\mathbf{n}_1,\mathbf{n}_2,t)}{dt} = \Omega \sum_{q=1}^{R} \left(\prod_{i=1,2} \prod_{\ell} \mathbb{E}^{-\hat{S}_{i,\ell;q}} - 1 \right) \hat{f}_q(\mathbf{n}_1/\Omega, \mathbf{n}_2/\Omega) P(\mathbf{n}_1,\mathbf{n}_2,t) \tag{9.3.11}$$

where $\mathbf{n}_j = \{n_{\ell,j}, \ell \in \mathscr{L}\}$, and

$$\hat{f}_q = f_r(u_{\ell,1}, u_{\ell,2}), \quad \hat{S}_{\ell,i;q} = S_{ir} \text{ for } q = (r,\ell)$$
$$\hat{f}_q = \frac{2\alpha_j}{z} u_{\ell,j}, \quad \hat{S}_{\ell,j;q} = -1, \hat{S}_{\ell',j;q} = 1 \text{ for } q = (j,\ell,\ell'), \quad \ell' \in \partial\ell.$$

The RD master equation (9.3.11) is clearly difficult to analyze. However, if Ω is sufficiently large, then we can carry out a linear noise approximation to obtain a Langevin equation for a multivariate OU process (Sect. 6.3). That is, we approximate the RD master equation by an FP equation and then linearize about the homogeneous steady state (u_1^*, u_2^*) by setting

$$\frac{n_{\ell,j}}{\Omega} = u_{\ell,j} = u_j^* + \frac{1}{\sqrt{\Omega}} v_{\ell,j},$$

with $u_1^* = a, u_2^* = b/a$ for the Brusselator model. This yields

$$\frac{dv_{\ell,j}(t)}{dt} = \sum_{j'} A_{jj'} v_{\ell,j'}(t) + \eta_{\ell,j}(t), \tag{9.3.12}$$

with white noise terms satisfying

$$\langle \eta_{\ell,j} \rangle = 0, \quad \langle \eta_{\ell,j} \eta_{\ell',j'} \rangle = D_{\ell,j;\ell',j'}.$$

Here

$$A_{jj'} v_{\ell,j'} = \sum_{r=1}^{p} S_{jr} \frac{\partial f_r(u_1^*, u_2^*)}{\partial u_{j'}^*} v_{\ell,j'} + \alpha_j \Delta v_{\ell,j} \delta_{j,j'} \tag{9.3.13}$$

and

$$D_{\ell,j;\ell',j'} = \delta_{\ell,\ell'} C_{jj'} + \frac{2\alpha_j}{z} (\Gamma^*)^j_{\ell\ell'} \delta_{j,j'} \tag{9.3.14}$$

with

$$\sum_{r=1}^{p} S_{jr} S_{j'r} f_r(u_1^*, u_2^*) = C_{jj'} \tag{9.3.15}$$

and

$$(\Gamma^*)^j_{\ell\ell} = 2z u_j^*, \quad (\Gamma^*)^j_{\ell\ell'} = -2u_j^* \text{ for all } \ell' \in \partial\ell. \tag{9.3.16}$$

Considerable insight into the behavior of the system can now be obtained by transforming to Fourier space [399, 429]. For simplicity, consider a 1D lattice with

periodic boundary conditions, $v_{\ell+N} = v_\ell$ for $\ell = 1,\ldots,N$, and set the lattice spacing $\Delta\ell = 1$. Introduce the discrete Fourier transforms (Box 2A)

$$V(k) = \sum_\ell \mathrm{e}^{-ik\ell} v_\ell, \quad v_\ell = \frac{1}{N}\sum_k \mathrm{e}^{ik\ell} V(k)$$

with $k = 2\pi m/N, m = 0,\ldots,N-1$. For a regular 1D lattice the Fourier transform of the discrete Laplacian operator is

$$\sum_\ell \mathrm{e}^{-ik\ell}\hat{\Delta} v_\ell = \sum_\ell \mathrm{e}^{-ik\ell}[v_{\ell+1} + v_{\ell-1} - 2v_\ell] = 2[\cos(k) - 1]V(k). \tag{9.3.17}$$

The discrete Fourier transform of the Langevin equation is

$$\frac{dV_j(k,t)}{dt} = \sum_{j'} A_{jj'}(k)V_{j'}(k,t) + \eta_j(k,t) \tag{9.3.18}$$

with

$$A_{jj'}(k) = \sum_{r=1}^{p} S_{jr}\frac{\partial f_r(u_1^*, u_2^*)}{\partial u_{j'}^*} + 2\alpha_j[\cos(k) - 1]\delta_{j,j'} \tag{9.3.19}$$

and

$$\langle \eta_j(k,t)\rangle = 0, \quad \langle \eta_j(k,t)\eta_{j'}(k',t')\rangle = \hat{D}_{jj'}(k,k')\delta(t-t').$$

Moreover, using the identity $\sum_\ell \mathrm{e}^{i(k-k')\ell} = N\delta_{k,k'}$, we have

$$\begin{aligned}
\hat{D}_{jj'}(k,k') &= \sum_{\ell,\ell'} \mathrm{e}^{-ik\ell}\mathrm{e}^{-ik'\ell'} D_{\ell,j;\ell',j'} \\
&= \sum_{\ell,\ell'} \mathrm{e}^{-ik\ell}\mathrm{e}^{-ik'\ell'}\left[\delta_{\ell,\ell'}C_{jj'} + \frac{2\alpha_j}{z}(\Gamma^*)^j_{\ell\ell'}\delta_{j,j'}\right] \\
&= N\delta_{k,-k'}\left(C_{jj'} + 4\alpha_j u_j^*[1-\cos(k)]\delta_{j,j'}\right) \\
&= N\delta_{k,-k'}C_{jj'}(k).
\end{aligned}$$

The factor of N can be eliminated by rescaling time t and $V_j(k)$ appropriately. Now Fourier transforming the Langevin equation with respect to time gives

$$\sum_l \Phi_{jl}(k,\omega)V_l(k,\omega) = \eta_j(k,\omega)$$

with

$$\Phi_{jl}(k,\omega) = -i\omega\delta_{j,l} - A_{jl}(k)$$

and

$$\langle \eta_j(k,\omega)\rangle = 0, \quad \langle \eta_j(k,\omega)\eta_{j'}(k',\omega')\rangle = \delta_{k,-k'}C_{jj'}(k)\delta(\omega+\omega').$$

Hence,

$$\begin{aligned}
&\langle V_i(k,\omega)V_i(k',\omega')\rangle \\
&= \left\langle \left[\sum_l \Phi_{il}^{-1}(k,\omega)\eta_l(k,\omega)\right]\left[\sum_j \Phi_{ij}^{-1}(k',\omega')\eta_j(k',\omega')\right]\right\rangle \\
&= \delta_{k,-k'}\delta(\omega+\omega')\sum_{l,j}\Phi_{il}^{-1}(k,\omega)C_{lj}(k)\Phi_{ij}^{-1}(-k,-\omega') \\
&= \delta_{k,-k'}\delta(\omega+\omega')\sum_{l,j}\Phi_{il}^{-1}(k,\omega)C_{lj}(k)(\Phi^{\dagger})_{ji}^{-1}(k,\omega).
\end{aligned}$$

Defining the power spectrum of the kth eigenmode by

$$\langle V_i(k,\omega)V_i(k',\omega')\rangle = S_i(k,\omega)\delta_{k,-k'}\delta(\omega+\omega'),$$

we deduce that

$$S_i(k,\omega) = \sum_{l,j}\Phi_{il}^{-1}(k,\omega)C_{lj}(k)(\Phi^{\dagger})_{ji}^{-1}(k,\omega). \tag{9.3.20}$$

Note that the above analysis applies to any two-species RD master equation of the form (9.3.11) and can be extended to multiple species. In the case of the Brusselator model system, whose deterministic RD equations are given by Eq. (9.3.6), we have

$$\sum_r S_{1r}f_r(u_{\ell,1},u_{\ell,2}) = a-(b+1)u_{\ell,1}+u_{\ell,1}^2u_{\ell,2}, \tag{9.3.21}$$

$$\sum_r S_{2r}f_r(u_{\ell,1},u_{\ell,2}) = bu_{\ell,1}-u_{\ell,1}^2u_{\ell,2} \tag{9.3.22}$$

and $\sum_{r=1}^4 S_{jr}S_{j'r}f_r(u_{\ell,1},u_{\ell,2}) = C_{jj'}^{\ell}$ with

$$\mathbf{C}^{\ell} = \begin{pmatrix} a+(b+1)u_{\ell,1}+u_{\ell,1}^2u_{\ell,2} & -bu_{\ell,1}-u_{\ell,1}^2u_{\ell,2} \\ -bu_{\ell,1}-u_{\ell,1}^2u_{\ell,2} & bu_{\ell,1}+u_{\ell,1}^2u_{\ell,2}\end{pmatrix}. \tag{9.3.23}$$

Carrying out the linear noise approximation and transforming to Fourier space then leads to the Langevin equation (9.3.18) with drift matrix

$$\mathbf{A}(k) = \begin{pmatrix} b-1+2\alpha_1[\cos(k)-1] & a^2 \\ -b & -a^2+2\alpha_2[\cos(k)-1]\end{pmatrix} \tag{9.3.24}$$

and diffusion matrix

$$\mathbf{C}(k) = \begin{pmatrix} 2(b+1)a+4\alpha_1u_1^*[1-\cos(k)] & -2ba \\ -2ba & 2ba+4\alpha_2u_2^*[1-\cos(k)]\end{pmatrix} \tag{9.3.25}$$

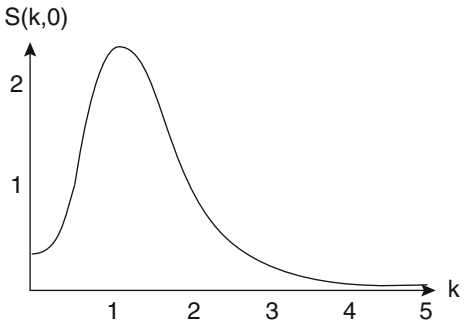


Fig. 9.12: Sketch of an example power spectrum $S(k,0)$ for the concentration of chemical X in a stochastic Brusselator model showing the peak in the spectrum at a nonzero wavenumber k

with $u_1^* = a, u_2^* = b/a$. For a wide range of RD models exhibiting a Turing instability in the deterministic limit, including the Brusselator model, one finds that when one plots $S_i(k,\omega)$ as a function of k, the power spectrum shows a peak at $k \neq 0, \omega = 0$, indicating the presence of a stochastic Turing pattern [47, 87, 88, 399, 429, 573, 689], see Fig. 9.12. Moreover, one can compute the region of parameter space for which the spectrum has a peak at nonzero k and demonstrate that these regions are significantly larger than the region over which a deterministic Turing pattern occurs. It also possible to extend the analysis to the case of Turing–Hopf bifurcations, where there is a peak in the power spectrum at $k \neq 0, \omega \neq 0$, which can result in either standing or traveling waves [48, 573]. The Turing–Hopf instability may be a bifurcation mechanism underlying the Min protein oscillations (Sect. 9.4).

One important extension of the above analysis would be to determine the effects of intrinsic noise on the selection and stability of patterns. One way to approach this would be to construct a stochastic version of the amplitude equations discussed in Sect. 9.2.2. However, it is not clear to what extent a diffusion approximation of the RD master equation is valid close to a bifurcation point where weakly nonlinear analysis can be applied. A related issue is to what extent symmetries of the underlying deterministic RD system persist in the presence of intrinsic noise.

9.4 Protein Self-Organization of Min Proteins in *E. coli*

The mechanism of cell division in bacteria differs significantly from eukaryotic cells; see Sect. 8.2. Bacterial cell division is initiated by the polymerization of the tubulin homolog FtsZ into the so-called Z-ring. In rod-shaped bacteria such as *E. coli*, formation of the Z-ring is usually restricted to the cell center, where it determines the site of cell division. The localization of the Z-ring is accurate to within 3 % of the cell length, resulting in two daughter cells of almost equal size. A major

process involved in the high precision of cell division is the regulatory system of Min proteins (see the review [396]). The Min systems consists of three proteins, MinC, MinD, and MinE. MinC inhibits Z-ring formation, whereas MinD and MinE act to confine MinC to the cell poles. A characteristic feature of Min protein dynamics is that the protein concentrations oscillate from pole to pole with a period of 1–2 min, which is much shorter than the cell cycle. Consequently, the time-averaged MinC concentration is maximized at the cell poles and minimized at the cell center, resulting in inhibition of Z-ring formation at the poles.

A more detailed picture of the mechanism underlying Min protein oscillations is shown in Fig. 9.13. MinC forms a complex with MinD and thus follows the spatiotemporal variation in MinD concentration, consistent with the finding that only MinD and MinE are essential for the occurrence of oscillations. The basic biochemical cycle is as follows: cytoplasmic MinD forms an ATPase MinD·ATP that binds cooperatively to the cell membrane, forming polymer filaments; MinE then binds to membrane MinD·ATP and stimulates ATP hydrolysis that causes MinD·ADP to be released from the membrane; the cytoplasmic MinD·ADP is then converted back to MinD·ATP and rebinds to the membrane. If MinD·ATP initially binds to a polar region, then it forms a cap that extends towards the cell center and is flanked by a ring of MinE known as the E-ring. The E-ring stimulates the hydrolysis of Min·ATP in its neighborhood, leading to the release of MinD·ADP and MinE, with the latter rebinding to the shrinking MinD·ATP cap. This results in movement of the E ring towards the pole in the wake of the shrinking MinD·ATP cap. Meanwhile, the released MinD·ATD reconverts to MinD·ATP and rebinds to the membrane at the opposite pole, where the MinE concentration is lowest, forming a new MinD·ATP cap. Once the first cap has disappeared, the released MinE rebinds to form an E-ring at the boundary of the new MinD·ATP cap. Iteration of this process underlies the observed Min protein oscillations.

A number of models have been developed that describe the interactions between MinD and MinE in terms of a system of reaction–diffusion (RD) equations [279, 284, 356, 432, 434]. All of the models can undergo a pattern forming instability of a homogeneous state, resulting in self-organized Min oscillations. However, the specifics of the molecular mechanism that generates the oscillations differ between the models, and have not yet been resolved conclusively by experiments. For example, Meinhardt and de Boer [434] consider a typical activator–inhibitor RD system with slow membrane diffusion and fast cytoplasmic diffusion (see Sect. 9.2); a crucial component of the model is protein synthesis and degradation. On the other hand, a crucial feature of the model by Howard et al. [275] is that cytoplasmic MinD recruits MinE to the membrane which in turn reduces the attachment rate of MinD. Subsequent models have emphasized the importance of MinD aggregation within the membrane and cooperative binding [284, 356, 432]. For the sake of illustration, consider the RD model of Huang et al. [284]. Let ρ_D, ρ_D^*, and ρ_E denote the concentrations of MinD·ATD, MinD·ATP, and MinE in the cytoplasm, and let ρ_d^*, ρ_{de} denote the concentrations of MinD·ATP and MinD·ATP·MinE complexes in the membrane. The cell is modeled as a cylinder of radius R and length L; typical values are $R = 0.5\,\mu\text{m}$ and $L = 5\,\mu\text{m}$. The various components of the model are as follows:

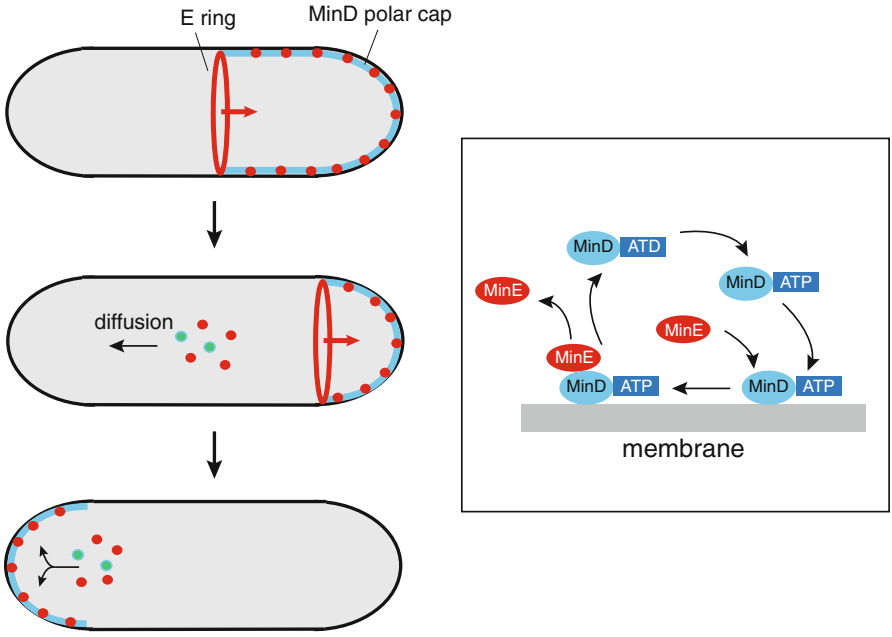


Fig. 9.13: Oscillatory patterns of Min protein system in *E. coli*. MinD·ATP (*blue*) binds to the membrane forming a polar cap. Min E (*red*) stimulates hydrolysis of MinD·ATP, which leads to protein release from the membrane. The polar cap shrinks and cytoplasmic MinD reverts to its ATP form and subsequently binds to the membrane at the opposite pole. *Inset* shows the cycle of biochemical reactions underlying the oscillations

1. Conversion of cytoplasmic MinD·ATD to MinD·ATP at a rate $\sigma_D \rho_D$
2. Hydrolysis of MinE-mediated membrane-bound MinD·ATP· at a rate $\sigma_d \rho_{de}$; MinE and MinD·ATD are then immediately released from the membrane.
3. The cooperative binding of cytoplasmic MinD·ATP at a rate

$$[k_D + k_d(\rho_{de} + \rho_d^*)]\rho_D^*(R),$$

   where $\rho_D^*(R)$ is the concentration of cytoplasmic MinD·ATP close to the membrane.
4. The binding of cytoplasmic MinE to membrane-bound MinD·ATP at a rate $\sigma_E \rho_d^* \rho_E$.
5. All cytoplasmic proteins have the same diffusion coefficient, which has the typical value $D = 2.5\,\mu\text{m}^2/\text{s}$, whereas membrane diffusion is assumed to be negligible.

The resulting system of RD equation is then

$$\frac{\partial \rho_D}{\partial t} = D\nabla^2 \rho_D - \sigma_D \rho_D + \delta(r-R)\sigma_d \rho_{de} \tag{9.4.1a}$$

$$\frac{\partial \rho_D^*}{\partial t} = D\nabla^2 \rho_D^* + \sigma_D \rho_D - \delta(r-R)[k_D + k_d(\rho_{de} + \rho_d^*)]\rho_D^* \tag{9.4.1b}$$

$$\frac{\partial \rho_E}{\partial t} = D\nabla^2 \rho_E + \delta(r-R)\sigma_d \rho_{de} - \delta(r-R)\sigma_E \rho_d^* \rho_E \tag{9.4.1c}$$

$$\frac{\partial \rho_d^*}{\partial t} = -\sigma_E \rho_d^* \rho_E(R) + [k_D + k_d(\rho_{de} + \rho_d^*)]\rho_D^*(R) \tag{9.4.1d}$$

$$\frac{\partial \rho_{de}}{\partial t} = -\sigma_d \rho_{de} + \sigma_E \rho_d^* \rho_E(R). \tag{9.4.1e}$$

Numerical simulations of the above system reveal oscillations with similar characteristics to those found in experiments, including the growth and shrinkage of alternating polar caps and the formation of an E-ring at the cell center. The reason that MinD can form a cap on the opposite pole is that it diffuses farther than MinE following release from the old polar cap, which is a consequence of the delay in converting MinD·ATD back to MinD·ATP. Once the original cap has disappeared, the newly released MinE proteins diffuse until they encounter the edge of the newly formed cap where they rapidly bind to form the E-ring.

The protein aggregation model of Kruse [356, 432] takes self-aggregation of MinD in the membrane rather than cooperative binding as a critical mechanism for Min oscillations. It also neglects details of ATP hydrolysis and assumes that the radial distribution of cytoplasmic proteins is uniform. The corresponding system of RD equations can be expressed in terms of the axial variable x as follows:

$$\frac{\partial \rho_D}{\partial t} = D\frac{\partial^2 \rho_D}{\partial x^2} + \sigma_d \rho_{de} - k_D(\rho_{\max} - \rho_{de} - \rho_d)\rho_D \tag{9.4.2a}$$

$$\frac{\partial \rho_E}{\partial t} = D\frac{\partial^2 \rho_E}{\partial x^2} + \sigma_d \rho_{de} - \sigma_E \rho_d \rho_E \tag{9.4.2b}$$

$$\frac{\partial \rho_d}{\partial t} = -\sigma_E \rho_d \rho_E + k_D(\rho_{\max} - \rho_{de} - \rho_d)\rho_D - \frac{\partial J_d}{\partial x} \tag{9.4.2c}$$

$$\frac{\partial \rho_{de}}{\partial t} = -\sigma_d \rho_{de} + \sigma_E \rho_d \rho_E. \tag{9.4.2d}$$

Here there is no distinction between ATP and ATD versions of MinD, and the binding rate of MinD simply depends on the product of the cytoplasmic MinD concentration ρ_D and the density of unoccupied membrane $\rho_{\max} - \rho_{de} - \rho_d$. The term $\partial_x J_d$ represents a membrane aggregation current,

$$J_d = -D_d \partial \rho_d + \rho_d(\rho_{\max} - \rho_{de} - \rho_d)\left[k_1 \partial_x \rho_d + k_2 \partial_x^3 \rho_d + \bar{k}_1 \partial_x \rho_{de} + \bar{k}_2 \partial_x^3 \rho_{de}\right], \tag{9.4.3}$$

where D_d is membrane diffusivity of MinD. (For simplicity, we drop a corresponding aggregation current for MinD·MinE complexes and set $\bar{k}_1 = \bar{k}_2 = 0$.) No-flux

boundary conditions are imposed at the ends $x = 0, L$ and the total protein numbers are conserved:

$$\int_0^L (\rho_D + \rho_d + \rho_{de})dx = LN_D, \quad \int_0^L (\rho_E + \rho_{de})dx = LN_E. \tag{9.4.4}$$

Following the analysis of [432], suppose that the diffusion of cytoplasmic proteins is much faster than the period of Min oscillations so that ρ_D and ρ_E can be approximated by uniform concentrations $\bar{\rho}_D$ and $\bar{\rho}_E$. Integrating equations (9.4.2a,b) with respect to x and using the conservation conditions leads to the pair of ODEs

$$\frac{d\bar{\rho}_D}{dt} = \sigma_d(N_D - \bar{\rho}_E) - k_D(\rho_{\max} - N_D + \bar{\rho}_D)\bar{\rho}_D \tag{9.4.5a}$$

$$\frac{d\bar{\rho}_E}{dt} = \sigma_d(N_D - \bar{\rho}_E) - \sigma_E(N_D - N_E - \bar{\rho}_D + \bar{\rho}_E)\bar{\rho}_E. \tag{9.4.5b}$$

It can be shown that there exists a unique fixed point in the domain $0 \le \bar{\rho}_D \le N_D$ and $0 \le \bar{\rho}_E \le N_E$. Substituting this fixed point into Eq. (9.4.2cd) leads to a pair of RD equations for membrane-bound Min proteins:

$$\frac{\partial \rho_d}{\partial t} = -\bar{\sigma}_E \rho_d + \bar{k}_D(\rho_{\max} - \rho_{de} - \rho_d) - \frac{\partial J_d}{\partial x} \tag{9.4.6a}$$

$$\frac{\partial \rho_{de}}{\partial t} = -\sigma_d \rho_{de} + \bar{\sigma}_E \rho_d, \tag{9.4.6b}$$

where $\bar{\sigma}_E = \sigma_E \bar{\rho}_E, \bar{k}_D = k_D \bar{\rho}_D$. Note the reaction terms are linear and determine the relaxation to a stationary state, whereas the current term is nonlinear and causes the pattern forming instability. This differs from classical RD equations such as the previous model, where transport is due to diffusion and instabilities arise from nonlinear reactions. Equation (9.4.6) have a homogeneous fixed point solution

$$\bar{\rho}_d = N_D - N_E - \bar{\rho}_D + \bar{\rho}_E, \quad \bar{\rho}_{de} = N_E - \bar{\rho}_E.$$

Linearizing about this fixed point, one can show that the resulting pair of linear PDEs have eigensolutions of the form

$$\rho_d - \bar{\rho}_d \sim \cos(\Omega t)\cos(q_c x), \quad \rho_{de} - \bar{\rho}_{de} \sim \cos(\Omega t + \phi)\cos(q_c x),$$

with corresponding eigenvalue $\lambda(q)$. The following results can be established (see Ex. 9.6): (i) The homogeneous fixed point is stable unless k_1 exceeds a critical value k_{1c}. (ii) The fastest growing (critical) eigenmode is an inhomogeneous stationary state ($\Omega = 0$) if $\sigma_d > \sigma_{d,c}$ and an inhomogeneous oscillatory pattern when $\sigma_d < \sigma_{d,c}$ for a critical parameter value $\sigma_{d,c}$. (iii) The critical wavenumber and frequency of the critical oscillatory mode are given by

$$q_c^4 = \frac{\sigma_D \bar{\rho}_D + \sigma_d + \sigma_E \bar{\rho}_E}{\bar{\rho}_d(\rho_{\max} - \bar{\rho}_d - \bar{\rho}_{de})k_2}, \quad \Omega_c^2 = \sigma_D \sigma_E \bar{\rho}_D \bar{\rho}_E - \sigma_d^2.$$

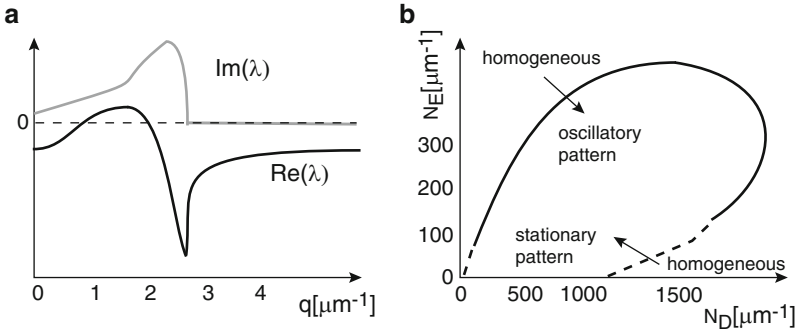


Fig. 9.14: Linear stability of the homogeneous state in aggregation current model of Min protein oscillations. (**a**) Sketch of example dispersion curves for the real and imaginary parts of the eigenvalue λ as a function of wavenumber q. There is a band of growing eigenmodes with nonzero wavenumber such that $\mathrm{I}m(\lambda) \neq 0$, indicating the formation of an oscillatory standing wave. (**b**) Sketch of stability diagram for the homogeneous fixed point as a function of the total densities of MinD and MinE. Crossing the solid (*dashed*) stability curve leads to an oscillatory (*stationary*) pattern. The *curves* are redrawn from [432] using the parameter values $\sigma_D = 4 \times 10^{-5}\,\mu\mathrm{ms}^{-1}, \sigma_E = 3 \times 10^{-4}\,\mu\mathrm{ms}^{-1}, \sigma_d = 0.04\,\mathrm{s}^{-1}, D_d = 0.06\,\mu\mathrm{m}^2\mathrm{s}^{-1}, \sigma_{\max} = 1{,}000\,\mu\mathrm{m}^{-1}, k_1 = 1.5 \times 10^{-6}\,\mu\mathrm{m}^4\mathrm{s}^{-1}, k_2 = 1.8 \times 10^{-7}\mu\mathrm{m}^6\mathrm{s}^{-1}, \bar{k}_1 = -1.2 \times 10^{-6}\,\mu\mathrm{m}^4\mathrm{s}^{-1}, \bar{k}_2 = 1.2 \times 10^{-10}\,\mu\mathrm{m}^6\mathrm{s}^{-1}$. In (**a**) $N_D = 900\,\mu\mathrm{m}^{-1}$ and $N_E = 350\,\mu\mathrm{m}^{-1}$

Example dispersion curves for the real and imaginary parts of $\lambda(q)$ together with a bifurcation diagram are sketched in Fig. 9.14.

As in the case of intracellular concentration gradients (Sect. 9.1), an important issue is the effect of intrinsic noise due to low protein copy numbers on Min oscillations. This has been investigated by Kruse et al. [184] in a combined experimental/computational study. The authors observed that for short cells (less than 2.5 μm), instead of exhibiting regular oscillations, MinD tended to shift stochastically from one cell half to the other. However, as the cell grew beyond around 2.5–3 μm, stochastic switching transitioned into regular Min oscillations, which became progressively more robust. Interestingly, the transition was accompanied by a qualitative change in the ensemble-averaged distribution $\rho(\tau)$ of residence times τ [184]. For small cells, the distribution appeared algebraic with $\rho(\tau) \sim \tau^{-\alpha}$ and $\alpha \approx 2.1$. (This is consistent with the expected exponential distribution of switching events for an individual cell, assuming a Gaussain variation in system parameters across cells.) On the other hand, for longer cells, $\rho(\tau)$ was sharply peaked at approximately half the period of oscillations (around 35 s), with a small tail of rare events in which

one or more oscillations were missed. An accompanying computational study of stochastic versions of several different models of Min Oscillations suggests that the aggregation model of Meacci and Kruse [432] best captures the stochastic switching observed experimentally [184]. One important issue is identifying the origin of the transition between stochastic switching and oscillations of MinD as the cell grows. It turns out that increasing cell length is not sufficient, assuming that the total number of proteins grows linearly with cell length. On the other hand, increasing the rate σ_E of MinE binding can trigger the transition.

9.5 Cell Polarization

Many cellular processes depend critically on the establishment and maintenance of polarized distributions of signaling proteins on the plasma membrane [300]. These include cell motility, neurite growth and differentiation, epithileal morphogenesis, embryogenesis, and stem cell differentiation. Cell polarization typically occurs in response to some external spatial cue such as a chemical gradient. A number of features of the stimulus response are shared by many different cell types, including amplification of spatial asymmetries, persistence of polarity when the triggering stimulus is removed, and sensitivity to new stimuli whereby a cell can reorient when the stimulus gradient is changed. In many cases, cell polarity can also occur spontaneously, in the absence of pre-existing spatial cues.

One of the most studied model systems of cell polarization is the budding yeast *Saccharomyces cerevisiae* [305, 598]. A yeast cell in the G1 phase of its life cycle (see Fig. 8.6) is spherical and grows isotropically. It then undergoes one of two fates: either it enters the mitotic phase of the life cycle and grows a bud, or it forms a mating projection (shmoo) towards a cell of the opposite mating type. Both processes involve some form of symmetry-breaking mechanism that switches the cell from isotropic growth to growth along a polarized axis (see Fig. 9.15). Under physiolog-

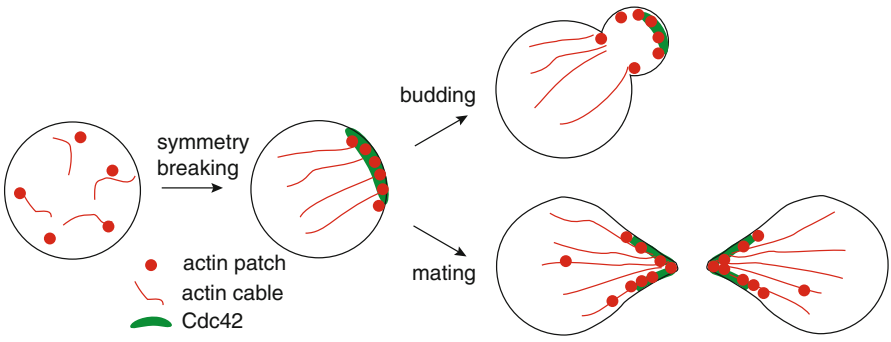


Fig. 9.15: Symmetry-breaking processes in the life cycle of budding yeast. See text for details. Redrawn from [598]

ical conditions, yeast cells polarize towards an environmental spatial asymmetry. This could be a pheromone gradient in the case of mating or a bud scar deposited on the cell surface from a previous division cycle. However, yeast cells can also undergo spontaneous cell polarization in a random orientation when external asymmetries are removed. For example, in the case of budding, induced cell mutations can eliminate the recognition of bud scars. Moreover, shmoo formation can occur in the presence of a uniform pheromone concentration. The observation that cells can break symmetry spontaneously suggests that polarization is a consequence of internal biochemical states. Experimental studies in yeast have shown that cell polarization involves a positive feedback mechanism that enables signaling molecules already localized on the plasma membrane to recruit more signaling molecules from the cytoplasm, resulting in a polarized distribution of surface molecules. The particular signaling molecule in budding yeast is the Rho GTPase Cdc42. (More complex cells involve interactions between several different GTPases; see below.) As with other Rho GTPases (see Box 9C), Cdc42 targets downstream affectors of the actin cytoskeleton. There are two main types of actin structure involved in the polarized growth of yeast cells: cables and patches. Actin patches consist of networks of branched actin filaments nucleated by the Arp2/3 complex at the plasma membrane, whereas actin cables consist of long, unbranched bundles of actin filaments. Myosin motors travel along the cables unidirectionally towards the actin barbed ends at the plasma membrane, transporting intracellular cargo such as vesicles, mRNA, and organelles. The patches act to recycle membrane-bound structures to the cytoplasm via endocytosis. During cell polarization, Cdc42-GTP positively regulates the nucleation of both types of actin structure, resulting in a polarized actin network, in which actin patches are concentrated near the site of asymmetric growth and cables are oriented towards the direction of growth. There are at least two independent but coordinated positive feedback mechanisms that can establish cell polarity [674]. One involves the reinforcement of spatial asymmetries by the directed transport of Cdc42 along the actin cytoskeleton to specific locations on the plasma membrane [374, 415], whereas the other involves an actin-independent pathway, in which Bem1, an adaptor protein with multiple binding sites, forms a complex with Cdc42 that enables recruitment of more Cdc42 to the plasma membrane. In the latter case, intrinsic noise plays an essential role in allowing positive feedback alone to account for spontaneous cell polarization [7, 302, 429]. An alternative possibility is that positive feedback is coupled with activity-dependent inhibition [300], resulting in a reaction–diffusion system that exhibits Turing pattern formation [231, 437] or bistability [200, 452, 581]. A universal feature of all of these models, which is also shared by other cell types, is that there exists at least one signaling molecule that can either be in an active membrane-bound state or an inactive cytosolic state, together with some form of feedback mediated by molecular interactions within the membrane or between the membrane and the cytosol [300].

Box 9C. Rho GTPase and cell polarization.

Rho GTPase is a class of signaling molecule that plays an important role in the polarization and migration of many different cell types [251]. Rho GTPases act as intracellular molecular switches that cycle between an active GTP-bound form in the membrane and an inactive GDP-bound form in the cytosol (see Fix. 9.16a). Guanine nucleotide exchange factors (RhoGEFs) facilitate the conversion from GDP-bound to GTP-bound form, whereas GTPase activating proteins (RhoGAPs) enhance GTP hydrolysis and are thus negative regulators. RhoGEFs and RhoGAPs are both regulated by upstream signals. A major downstream target of the Rho GTPase signaling pathways is the actin cytoskeleton [251, 400, 468]. A variety of actin accessory proteins mediate the different components of actin dynamics within a cell and can be activated by the various signaling pathways associated with cell polarization. We describe a few of the major players; see [468] for more details. First, the *actin related proteins 2 and 3* (Arp2/3) *complex* stimulates actin polymerization by creating new nucleation cores. The Arp2/3 complex is activated by members of the *Wiskott–Aldrich syndrome protein* (WASP) family protein (WAVE) complex, which localizes to lamellipodia (mesh-like actin sheets) where it facilitates actin polymerization.

The WAVE complex is activated by the small GTPase *Rac1*, which modulates the actin cytoskeleton dynamics by controlling the formation of lamellipodia. The formation of filopodia (packed actin bundles), on the other hand, is regulated by another member of the small GTPases, called *Cdc42*. A second important accessory protein is *cofilin*, a member of the actin depolymerizing factor (ADF)/cofilin family, which modifies actin dynamics by increased severing and depolymerization of actin filaments via its binding to the non-barbed (pointed) ends. Cofilin is inhibited when phosphorylated by LIM-kinase, which is itself activated by p21 activated kinases (PAKs). Since the latter is a downstream target of Rac1 and Cdc42, it follows that Rac1 and Cdc42 inhibit actin depolymerization by downregulating cofilin. At first sight, one would expect the reduction of cofilin to counter the affects of Arp2/3. However, the extent to which cofilin enhances or reduces cell protrusion depends on the spatial and temporal scale over which it operates [136]. It turns out that the overall effect of Rac1 and Cdc42 is to increase actin dynamics, thus promoting cytoskeletal growth. This is opposed by the action of a third type of GTPase known as *RhoA*, which tends to stabilize the actin network by activating ROCK-kinase, which promotes *profilin* (an actin binding protein that catalyzes polymerization) and suppresses cofilin. A summary of the basic signaling pathways is given in Fig. 9.16b.

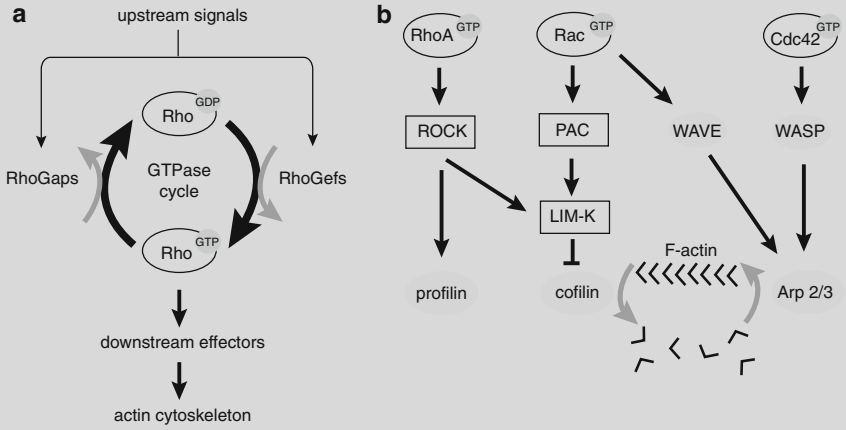


Fig. 9.16: (**a**) The Rho GTPase cycle. (**b**) Simplified signaling pathways from GTPases RhoA, Rac, and Cdc42 to actin accessory proteins cofilin and Arp 2/3 that regulate actin polymerization in the cytoskeleton (Redrawn from Luo [400])

Probably the most striking example of a polarized cell is the neuron, due to its compartmentalization into a thin, long axon and several shorter, tapered dendrites. Experimental studies of neuronal polarization have mainly been performed on dissociated, embryonic cortical and hippocampal neurons or on postnatal cerebellar granule neurons. Such studies have identified three basic stages of polarization [12, 468]; see Fig. 9.17. Cultured neurons initially attach to their substrate as round spheres surrounded by actin-rich structures including lamellipodia and filopodia (stage 1). Lamellipodia then coalesce to form *growth cones*, followed by the establishment of several short processes, called *neurites* (stage 2). Eventually one of the neurites starts to grow more rapidly to become the axon (stage 3), while the other neurites remain short and develop into dendrites at later stages of maturation. The growth cone at the mobile tip of an elongating neurite or axon contains microtubules within a central domain (C-domain) and actin filaments within the peripheral domain (P-domain) (see Fig. 9.18). The microtubules provide the structural backbone of the shaft and a substrate for intracellular transport to the growth cone. They polymerize with their growing ends pointed towards the leading edge of the growth cone. Actin filaments within the P-domain form filopodia and lamellipodia that shape and direct the motility of the growth cone. In both structures, the actin filaments face with their barbed (growing) ends towards the plasma membrane. Polymerization of actin filaments towards the leading edge causes the extension and protrusion of the growth cone. This creates a force that pushes the actin network and the tightly linked plasma membrane backward (retrograde flow), and hinders the invasion of the microtubules into the P-domain. The retrograde flow is also enhanced by the action of myosin molecular motors, which drag the actin cytoskeleton back towards the C-domain where actin filaments depolymerize at their pointed ends. If there is

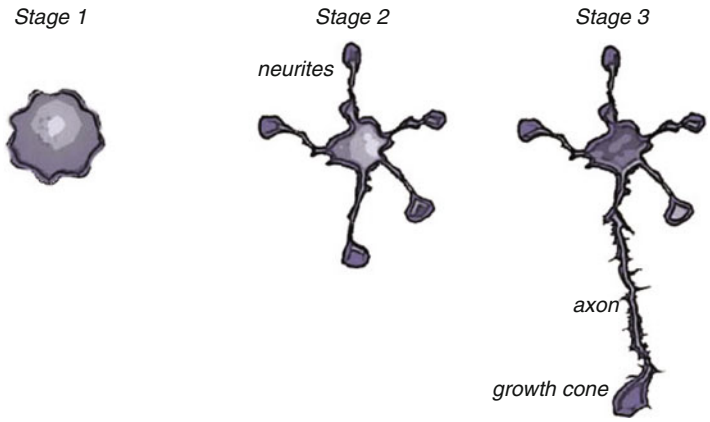


Fig. 9.17: Stages of neuronal polarization. A neuron attaches to the substrate as a round sphere surrounded by actin-rich lamellipodia and filopodia (stage 1). Growth cone formation by the consolidation of lamellipodia leads to the establishment of several neurites (stage 2). One neurite starts to elongate rapidly and forms the axon (stage 3)

a balance between actin polymerization in the P-domain and retrograde flow, then there is no elongation. However, signals from surface adhesion receptors bound to a substrate can suppress the retrograde flow of actin filaments, shifting the balance towards polymerization-driven forward motion that involves both actin filaments and microtubules.

The growth cone of an axon can itself exhibit a form of polarization. During neural development, the growth cone has to respond accurately to extracellular chemical gradients that direct its growth. One such gradient activates *gamma-aminobutyric acid* (GABA) receptors in the plasma membrane, which then redistribute themselves asymmetrically towards the gradient source. SPT experiments have shown how this

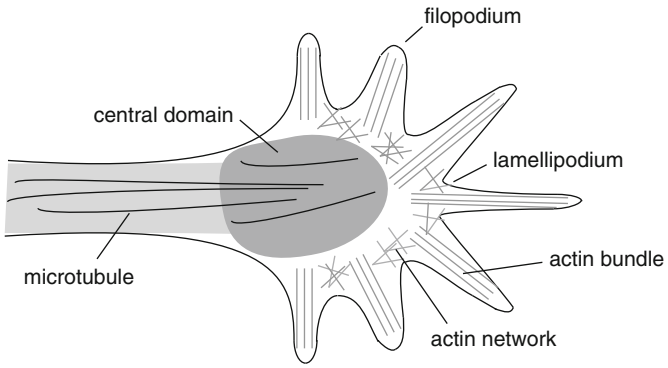


Fig. 9.18: Schematic diagram of growth cone showing cytoskeletal structures

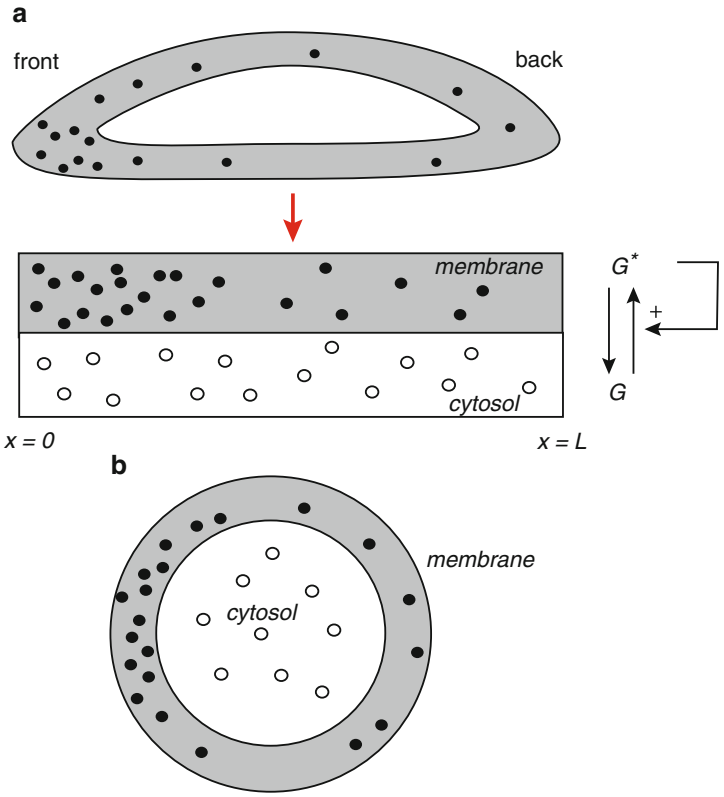


Fig. 9.19: Schematic diagram showing spatial distribution of a membrane-bound active signaling molecule G^* and a cytosolic inactive form G. (**a**) A 1D bidomain model of a flattened eukaryotic cell adhering to a substrate with no-flux boundary conditions at either end. (**b**) An idealized 2D or 3D cell with a spatially uniform interior and a polarized perimeter corresponding to the cell membrane

redistribution involves interactions between the GABA receptors and microtubules in the growth cone, which results in a reorientation of the microtubules towards the gradient source and subsequent steering of the growth cone [59]. Interestingly, in contrast to the interactions between Cdc42 and actin in budding yeast, one does not observe spontaneous polarization of the growth cone in the absence of an external spatial cue. This will be explored further in Sect. 9.5.3.

Another important example of cell polarization occurs in a variety of motile eukaryotic cells that undergo directed motion in response to external spatial signals—eukaryotic chemotaxis [271, 384, 417]. (See Sect. 5.3 for a corresponding discussion of bacterial chemotaxis.) Examples include mammalian neutrophils (white blood cells), fibroblasts (connective tissue mammalian cells responsible for wound healing), and keratocytes (fast moving epithelial cells from scales of fish). Prior to initiating movement, a given cell polarizes according to directional cues

in the environment, forming nascent "front" and "back" regions; see Fig. 9.19a. Protrusion of the front is driven by the assembly of actin cytoskeleton, whereas myosin motors at the back contract and pull up the rear. The polarization of the cell is regulated by Rho GTPases and phosphoinositides (PIs), with Rac, Cdc42, and PIP3 localized at the front and Rho localized at the back. (PIs are lipids that play an important role in regulating vesicular trafficking and actin polymerization.) Although different motile eukaryotic cells utilize much of the same molecular machinery, they can exhibit significantly different forms of behavior [300]. For example, neutrophils can sense very small gradients over a large range of concentrations, polarize very quickly (in less than a minute), and do not spontaneously polarize in the absence of chemoattractant. On the other hand, fibroblasts exhibit much less sensitivity to concentration gradients, polarize and move more slowly, and spontaneously polarize after being put on an adhesive substrate. Finally, keratocytes polarize on a similar time scale to neutrophils, spontaneously polarize after being detached from surrounding cells, and react to mechanical rather than chemical stimuli. There is one further important difference between models of motile eukaryotic cells and other examples of cell polarization such as budding yeast, which is illustrated in Fig. 9.19. Polarized eukaryotic cells such as fibroblasts and keratocytes tend to have a flattened shape when they adhere to a surface. If one imagines taking a cross-sectional slice along the polarization axis, one can treat the system as an effective 1D bidomain model with no-flux boundary conditions at either end. Both membrane-bound and cytosolic molecules diffuse along the polar axis, with a high concentration of membrane-bound molecules at the front and a low concentration at the back; see Fig. 9.19a. On the other hand, in the case of yeast, one treats the interior of the cell as spatially uniform and determines the spatial distribution of chemicals along the cell perimeter, which in the case of a simplified two-dimensional cell consists of a circle. Cell polarization would then correspond to a localized increase in concentration somewhere on the circle. In the remainder of this section, we describe in some detail several models of cell polarization in budding yeast: actin-independent positive feedback (Sects. 9.5.1 and 9.5.2), actin-dependent positive feedback (Sect. 9.5.3), and a Turing-based model of activation/inhibition (Sect. 9.5.4). Finally, in Sect. 9.5.5 we briefly describe a wave-pinning model of polarization in motile eukaryotic cells.

9.5.1 Positive Feedback Model: Deterministic Analysis

We begin by considering a mechanism for cell polarization based on positive feedback alone, which has been developed by Altschuler and collaborators [7, 300]. The basic framework is as follows: a signaling molecule such as Cdc42 transitions between inactive (cytosolic) and active (membrane-bound) states; activated molecules laterally diffuse along the membrane, recruiting inactive molecules within the cytosol to their membrane locations. In the case of a large number of signaling molecules N, the system converges to a spatially uniform steady state and

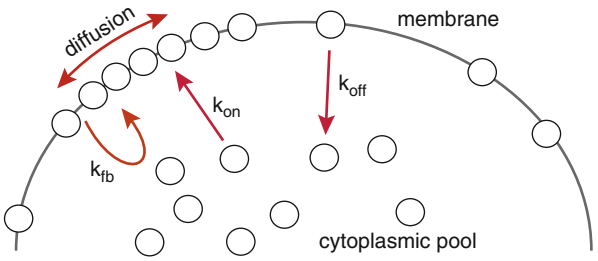


Fig. 9.20: Positive feedback model of cell polarization. Redrawn from [415]

localized clusters cannot form. However, cell polarity emerges for intermediate values of N due to the effects of intrinsic noise fluctuations. Although stochasticity and diffusion ultimately disperses the cluster, polarization is recurrent in the sense that there is a repeating sequence of cluster formation and dispersal. The basic assumptions of the model are as follows [7, 302] (see Fig. 9.20):

1. *Diffusion.* Molecules in the cell membrane diffuse via Brownian motion with diffusion coefficient D, while cytosolic molecules are uniformly distributed due to a much faster rate of diffusion.
2. *Mass conservation.* The total number of signaling molecules $N = n_c + n_m$ is fixed, where n_c (n_m) is the number of cytosolic (membrane-bound) molecules.
3. *Mass-action kinetics.* There are three mechanisms for transitioning between the active and inactive states: spontaneous dissociation from the membrane at a rate $k_{\rm off}$, spontaneous association at a rate $k_{\rm on}$, and recruitment via positive feedback at a rate $k_{\rm fb}$.

We begin by analyzing the deterministic version of the model, in which the number N of signaling molecules is sufficiently large so that fluctuations can be ignored; see also the supplementary material of [302]. For the moment, we also neglect the effects of diffusion, and consider changes in the total number of molecules n_c in the cytosol, whose dynamics can be described by deterministic kinetic equations based on the law of mass action:

$$\frac{dn_c}{dt} = K_{\rm off}(N - n_c) - K_{\rm on} n_c - K_{\rm fb} n_c (N - n_c). \tag{9.5.1}$$

The modified reaction rates are

$$K_{\rm off} = k_{\rm off}, \quad K_{\rm on} = \frac{V_{\rm on}}{V} k_{\rm on}, \quad K_{\rm fb} = \frac{V_{\rm fb}}{V} k_{\rm fb}.$$

The volume fractions take into account the fact that molecules are only spontaneously activated if they are within a region of volume $V_{\rm on}$ near the cell membrane, and additional molecules are only recruited via feedback if they are within a region of volume $V_{\rm fb}$ around a membrane-bound molecule. Here V is the total cell volume. Equation (9.5.1) can be rewritten in the more suggestive form

$$\frac{dn_c}{dt} = K_{\text{fb}}(n_c - N)(n_c - N_c) - K_{\text{on}} n_c, \quad N_c = \frac{K_{\text{off}}}{K_{\text{fb}}}. \tag{9.5.2}$$

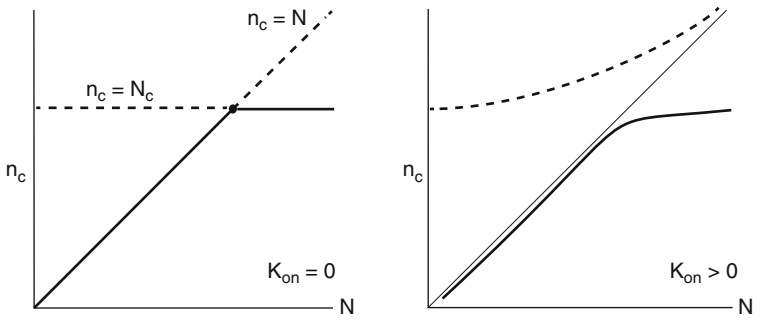


Fig. 9.21: Perfect (K_{on}) and imperfect ($K_{\text{on}} > 0$) transcritical bifurcations. In both cases the number of molecules n_c in the cytosol starts to significantly decrease the maximum N around the point $n_c = N_c$

If there are no spontaneous association events ($K_{\text{on}} = 0$), then we have a quadratic first-order ODE

$$\frac{dn_c}{dt} = K_{\text{fb}}(n_c - N)(n_c - N_c),$$

with a pair of fixed points at $n_c = N$ and $n_c = N_c$. If $N < N_c$, then the fixed point $n_c = N$ is stable and $n_c = N_c$ is unstable, whereas the converse holds when $N > N_c$. (Strictly speaking, the fixed point $n_c = N_c > N$ only makes sense mathematically.) The ODE is said to undergo a transcritical bifurcation at the critical molecular number $N = N_c$ since a pair of fixed points exchange their stability properties; (see Fig. 9.21). From a biological perspective, if the stable state is $n_c = N$, then all of the molecules are in the cytosol and no membrane clustering can occur. On the other hand, if $n_c = N_c$ then there are N_c molecules in the cytosol and $N - n_c$ in the cell membrane. It follows that a necessary condition for cell polarization to occur is that $N > N_c$, which means that N cannot be too small. If spontaneous activation events are now allowed ($K_{\text{on}} > 0$), then the above picture still approximately holds provided that $K_{\text{on}} \ll K_{\text{off}}$. The system now undergoes an imperfect transcritical bifurcation; although there exists a unique stable branch rather than an exchange of symmetries, there is still a switch in behavior around $n_c = N_c$.

The next step is to introduce diffusion into the deterministic mass-action model. Let $u(\mathbf{x},t)$ denote the density of signaling molecules in the cell membrane which forms a bounded 2D domain Σ. The density evolves according to the reaction-diffusion (RD) equation

$$\frac{\partial u}{\partial t} = D\nabla^2 u + \frac{K_{\text{on}}}{|\Sigma|} n_c + K_{\text{fb}} n_c u - K_{\text{off}} u, \tag{9.5.3}$$

where ∇^2 is the Laplace operator on Σ (or the more general Laplace–Beltrami operator in the case of a curved surface). This is supplemented by the Neumann boundary condition

$$\nabla u(\mathbf{x},t)\cdot\mathbf{n}=0 \text{ for all } \mathbf{x}\in\partial\Sigma,$$

where $\mathbf{n}$ is the unit normal on the membrane boundary. The second term on the right-hand side of Eq. (9.5.3) is divided by the area $|\Sigma|$ of the membrane in order to be consistent with the definition of u. Conservation of mass then requires that

$$n_c(t)+\int_\Sigma u(\mathbf{x},t)d\mathbf{x}=N.$$

Differentiating both sides and using Eq. (9.5.3) together with the boundary condition gives

$$\frac{dn_c}{dt}=-\frac{d}{dt}\int_\Sigma u(\mathbf{x},t)d\mathbf{x}=-K_{\rm on}n_c-K_{\rm fb}(N-n_c)n_c+K_{\rm off}(N-n_c). \tag{9.5.4}$$

We now show that the given linear RD equation cannot support any spatial patterning. Introduce the average density

$$\bar{u}(t)=\frac{1}{|\Sigma|}\int_\Sigma u(\mathbf{x},t)d\mathbf{x},$$

and the mean-square deviation from the average

$$\sigma^2(t)=\frac{1}{|\Sigma|}\int_\Sigma\left(\frac{u(\mathbf{x},t)-\bar{u}(t)}{\bar{u}(t)}\right)^2 d\mathbf{x}.$$

We will show that $\sigma(t)^2$ decays exponentially in time, which means that any spatial inhomogeneities are smoothed out by diffusion and hence cell polarization cannot occur. Defining $v(\mathbf{x},t)=[u(\mathbf{x},t)-\bar{u}(t)]/\bar{u}(t)$, one finds from Eqs. (9.5.3) and (9.5.4) that

$$\frac{\partial v}{\partial t}=D\nabla^2 v-\frac{K_{\rm on}n_c(t)}{|\Sigma|\bar{u}(t)}v.$$

It follows that

$$\begin{aligned}\frac{d}{dt}\int_\Sigma v^2 d\mathbf{x}&=2\int_\Sigma vv_t d\mathbf{x}\\&=2\int_\Sigma\left[vD\nabla^2 v-\frac{K_{\rm on}n_c(t)}{|\Sigma|\bar{u}(t)}v^2\right]d\mathbf{x}\\&\le 2D\int_\Sigma vD\nabla^2 v=-2D\int_\Sigma|\nabla v|^2 d\mathbf{x}.\end{aligned}$$

Finally, applying Poincare's inequality, Eq. (9.5.6) of Box 9D,

$$\frac{d}{dt}\int_\Sigma v^2 d\mathbf{x} \le -2D\int_\Sigma |\nabla v|^2 d\mathbf{x} \le -2\mu_1 D\int_\Sigma v^2 d\mathbf{x}$$

and noting that $\sigma(t)^2 = |\Sigma|^{-1}\int v^2 d\mathbf{x}$, we have

$$\frac{d\sigma^2}{dt} \le -2\mu_1 D\sigma^2, \tag{9.5.5}$$

where μ_1 is the first nonzero eigenvalue of the Laplacian. In conclusion, $\sigma(t)^2 \le \mathrm{e}^{-2\mu_1 Dt} \to 0$ as $t \to \infty$.

Box 9D. Poincare's inequality.

Consider the Laplacian operator ∇^2 acting on a bounded domain Σ with the Neumann boundary condition $\nabla\phi\cdot\mathbf{n} = 0$, where $\mathbf{n}$ is the unit normal to the boundary $\partial\Sigma$. The operator has a complete set of orthonormal eigenfunctions ϕ_k satisfying the equation

$$\nabla^2\phi_k + \mu_k\phi_k = 0,$$

with

$$\int_\Sigma \phi_k(\mathbf{x})\phi_l(\mathbf{x})d\mathbf{x} = \delta_{l,k}.$$

The eigenvalues μ_k are ordered such that

$$0 = \mu_0 < \mu_1 \le \mu_2 \le \mu_3 \ldots.$$

Suppose that $v(\mathbf{x})$ is a function in Σ that satisfies the Neumann boundary condition on $\partial\Sigma$ and $\int_\Sigma v(\mathbf{x})d\mathbf{x} = 0$. We can then expand v in terms of the generalized Fourier series

$$v(\mathbf{x}) = \sum_{k\ge 1} a_k\phi_k(\mathbf{x}), \quad a_k = \int_\Sigma v(\mathbf{x})\phi_k(\mathbf{x})d\mathbf{x}.$$

Note that $\phi_0(\mathbf{x}) =$ constant which means that $a_0 = 0$. The following result then holds:

$$\begin{aligned}\int_\Sigma |\nabla v|^2 d\mathbf{x} &= \int_\Sigma \left[\sum_{k\geq 0} a_k \nabla \phi_k(\mathbf{x})\right] \cdot \left[\sum_{l\geq 0} a_l \nabla \phi_l(\mathbf{x})\right] \\ &= \sum_{k,l\geq 0} a_k a_l \int_\Sigma \nabla \phi_k(\mathbf{x}) \cdot \nabla \phi_l(\mathbf{x}) d\mathbf{x} \\ &= \sum_{k,l\geq 0} a_k a_l \int_\Sigma \left[\nabla \cdot (\phi_k \nabla \phi_l) - \phi_k \nabla^2 \phi_l\right] d\mathbf{x} \\ &= \sum_{k,l\geq 0} a_k a_l \left[\int_{\partial\Sigma} \phi_k \nabla \phi_l \cdot \mathbf{n} d\sigma + \int_\Sigma \phi_k(\mathbf{x}) \mu_l \phi_l(\mathbf{x})\right] d\mathbf{x} \\ &= \sum_{k\geq 1} a_k^2 \mu_k \geq \mu_1 \sum_{k\geq 1} a_k^2.\end{aligned}$$

We have used the divergence theorem, the eigenvalue equation, and the Neumann boundary condition. Using a similar analysis, it is straightforward to show that

$$\int_\Sigma v(\mathbf{x})^2 d\mathbf{x} = \sum_{k\geq 1} a_k^2.$$

We thus obtain Poincare's inequality

$$\int_\Sigma v(\mathbf{x})^2 d\mathbf{x} \leq \frac{1}{\mu_1} \int_\Sigma |\nabla v(\mathbf{x})|^2 d\mathbf{x}. \tag{9.5.6}$$

9.5.2 *Positive Feedback Model: Stochastic Analysis*

Although the deterministic model does not support spontaneous cell polarization, numerical simulations of a stochastic version of the model reveal that a localized aggregate of membrane-bound molecules can form and persist for physiologically reasonable time periods, before ultimately dispersing due to the effects of diffusion [7, 302]. This stochastic-based effect has also been explored in more detailed simulations [373] and in a rigorous mathematical study [247]. Here we will describe a recent stochastic analysis based on a spatially discrete master equation [429]. This extends the Fokker–Planck approximation of chemical master equations to include the effects of diffusion, as outlined in Sect. 9.3. First, consider a stochastic extension of the diffusion-free model that takes into account finite number fluctuations using a birth–death master equation [302]. This takes the form of Eq. (3.2.2):

$$\begin{aligned}\frac{d}{dt} P(n,t) = \omega_+(n-1)P(n-1,t) + \omega_-(n+1)P(n+1,t) \\ - [\omega_+(n) + \omega_-(n)]P(n,t),\end{aligned} \tag{9.5.7}$$

where $P(n,t)$ is the probability that there are $n_c(t) = n$ cytosolic molecules at time t and the transition rates are

$$\omega_+(n) = (N-n)K_{\text{off}}, \quad \omega_-(n) = (N-n)nK_{\text{fb}} + nK_{\text{on}}. \tag{9.5.8}$$

Following the analysis of two-state ion channels in Sect. 3.2, we know that the steady-state solution is given by Eq. (3.2.6):

$$P_s(n) = P_s(0)\prod_{m=1}^{n}\frac{\omega_+(m-1)}{\omega_-(m)}, \tag{9.5.9}$$

with

$$P_s(0) = \left(1 + \sum_{n=1}^{N}\prod_{m=1}^{n}\frac{\omega_+(m-1)}{\omega_-(m)}\right)^{-1}.$$

Substituting the specific form of $\omega_\pm$ for the cell polarization model, we find that for $n < N$ (see Ex. 9.7)

$$\frac{P_s(n)}{P_s(0)} = \frac{N}{N-n}\frac{N_c^n}{n!}\prod_{m=1}^{n}\left[1+\frac{\gamma}{N-m}\right]^{-1}, \quad \gamma = \frac{K_{\text{on}}}{K_{\text{fb}}}. \tag{9.5.10}$$

Moreover, if $\gamma \ll 1$, then

$$\frac{P_s(n)}{P_s(0)} \approx \left(\frac{N}{N-n}\right)^{1-\gamma}\frac{N_c^n}{n!}, \quad \frac{P_s(N)}{P_s(0)} = \frac{1}{\gamma}\frac{N_c^n}{n!}N^{1-\gamma}.$$

From these latter expressions, it can be shown that for $N < N_c$, the stationary distribution is peaked around $n = N$, whereas for $N > N_c$ it is essentially a Poisson distribution peaked around $n = N_c$ with width $\sqrt{N_c}$. Interestingly, close to the bifurcation point of the deterministic system and for a certain range of values of γ, the probability distribution is bimodal with peaks around N_c and N [302].

We can incorporate diffusion into the master equation framework along similar lines to the analysis of stochastic pattern formation presented in Sect. 9.3, which was adapted to the particular problem of cell polarity by McKane et al. [429]. Thus, we partition the membrane Σ into small domains with centers at N discrete lattice points ℓ. For simplicity, it is assumed that the domain centers are distributed on a regular d-dimensional lattice ($d = 1,2$) with lattice spacing $\Delta\ell$. Let m_ℓ denote the number of signaling molecules in the domain with center at ℓ and introduce the local densities $u_\ell = m_\ell/\Omega$, where Ω is the total number of molecules in the cell. Similarly, denote the density of molecules in the cytoplasm by $v = n_c/\Omega$. Finally, assume that the feedback rate can be scaled as $K_{\text{fb}} \to K_{\text{fb}}/\Omega$ with all rate constants now independent of Ω. A spatially discrete version of the RD equation (9.5.3) then takes the form

$$\frac{du_\ell}{dt} = K_{\text{on}}v + K_{\text{fb}}u_\ell v - K_{\text{off}}u_\ell + \alpha\hat{\Delta}u_\ell, \tag{9.5.11}$$

where $\hat{\Delta}$ is the discrete Laplacian Eq. (9.3.7). Equation (9.5.11) is supplemented by the number conservation condition

$$v + \sum_{\ell} u_{\ell} = 1. \tag{9.5.12}$$

The continuum RD equation (9.5.3) is recovered in the continuum limit $\Delta\ell \to 0$ and $\alpha \to \infty$ with $2\alpha(\Delta\ell)^2 = D$ fixed. For this particular RD model, there are N "chemical species" labeled by ℓ (since n_c can be eliminated using the conservation condition), zN hopping reactions, and $2N$ reactions between the membrane and cytoplasm so that the total number of reactions is $R = (z+2)N$: For each lattice site ℓ,

$$m_{\ell} \xrightarrow{T_{\ell}^{+}} m_{\ell} + 1, \quad m_{\ell} \xrightarrow{T_{\ell}^{-}} m_{\ell} - 1$$

and we have the hopping reactions

$$(m_{\ell}, m_{\ell'}) \xrightarrow{T_{\ell\ell'}} (m_{\ell} - 1, m_{\ell'} + 1), \quad \ell' \in \partial\ell.$$

The corresponding transition rates are

$$T_{\ell}^{+} = \left[K_{\rm on} + \frac{K_{\rm fb}}{\Omega} m_{\ell} \right] (\Omega - \sum_{\ell} m_{\ell}), \quad T_{\ell}^{-} = K_{\rm off} m_{\ell}, \quad T_{\ell\ell'} = \frac{2\alpha}{z} m_{\ell}.$$

It is now straightforward to read off the functions f_a and stochiometric coefficients $S_{a\ell}$ for all the reactions labeled by a and then carry out the steps to reduce the associated master equation (6.3.18) to the FP equation (6.3.20). The corresponding Langevin equation is

$$\frac{dU_{\ell}(t)}{dt} = V_{\ell}(\mathbf{U}) U_{\ell}(t) + \Omega^{-1/2} \eta_{\ell}(t), \tag{9.5.13}$$

with white noise terms satisfying

$$\langle \eta_{\ell} \rangle = 0, \quad \langle \eta_{\ell} \eta_{\ell'} \rangle = D_{\ell\ell'},$$

such that

$$V_{\ell}(\mathbf{u}) = [K_{\rm on} + K_{\rm fb} u_{\ell}] (1 - \sum_{\ell} u_{\ell}) - K_{\rm off} u_{\ell} + \alpha \hat{\Delta} u_{\ell} \tag{9.5.14}$$

and

$$D_{\ell\ell'}(\mathbf{u}) = \left[[K_{\rm on} + K_{\rm fb} u_{\ell}] (1 - \sum_{\ell} u_{\ell}) + K_{\rm off} u_{\ell} + O(\alpha) \right] \delta_{\ell,\ell'}. \tag{9.5.15}$$

Note that contributions to the noise from hopping reactions will be negligible when α is small, since there is also the additional small factor $1/\Omega^{1/2}$ in Eq. (9.5.13). In the following, we drop the $O(\alpha)$ contributions to the diffusion matrix and set $K_{\rm on} = 0$.

As in the case of stochastic pattern formation (Sect. 9.3), considerable insight into the behavior of the system can be obtained by transforming to Fourier space [399, 429]. Again we consider a 1D lattice with periodic boundary conditions, $u_{\ell+N\Delta\ell} = u_\ell$ for all ℓ and lattice spacing $\Delta\ell$. Introduce the discrete Fourier transforms

$$\hat{u}(k) = \Delta\ell \sum_\ell \mathrm{e}^{-ik\ell} u_\ell, \quad u_\ell = \frac{1}{N\Delta\ell} \sum_k \mathrm{e}^{ik\ell} \hat{u}(k)$$

with $k = 2\pi m/(N\Delta\ell), m = 0, \ldots, N-1$. For a regular 1D lattice the Fourier transform of the discrete Laplacian operator is given by Eq. (9.3.17). Combining this with the observation that $\sum_\ell u_\ell = (\Delta\ell)^{-1}\hat{u}(0)$, the Fourier transform of the deterministic equation (9.5.11) is

$$\frac{d\hat{u}(k)}{dt} = \left[K_{\mathrm{fb}}(1 - (\Delta\ell)^{-1}\hat{u}(0)) - K_{\mathrm{off}} + 2\alpha(\cos(k\Delta\ell) - 1)\right]\hat{u}(k). \tag{9.5.16}$$

Recall that for $K_{\mathrm{fb}} > K_{\mathrm{off}}$ the deterministic equation (9.5.11) has a homogeneous fixed point $u_\ell = u^* = 1 - K_{\mathrm{off}}/K_{\mathrm{fb}}$ for all ℓ. In Fourier space this corresponds to the solution $\hat{u}(k) = \delta_{k,0}(\Delta\ell)u^*$. Linearizing Eq. (9.5.16) about the fixed point yields a diagonal Jacobian with eigenvalues

$$\lambda_0 = K_{\mathrm{off}} - K_{\mathrm{fb}}, \quad \lambda_k = 2\alpha[\cos(k\Delta\ell) - 1] \text{ for } k \neq 0. \tag{9.5.17}$$

We deduce that the homogeneous fixed point is stable, which appears to preclude any polarization. However, as highlighted in [429], if α is small, then spatially varying modes ($k \neq 0$) decay slowly so that a random perturbation combined with nonlinearities in the system could set up a polarized state that persists for long times. It turns out that Fourier analysis of the Langevin equation (9.5.13) (see Box 9E) yields the result

$$\langle|\hat{U}(k,t)|^2\rangle \approx \frac{(u^*\Delta\ell)^2}{1 + \Gamma^2 k^2}, \quad \Gamma = \sqrt{2\alpha\Omega u^* \Delta\ell^2/K_{\mathrm{off}}}. \tag{9.5.18}$$

The significance of the above result for cell polarization can be understood by taking averages of the identity

$$\Delta\ell \sum_{\ell'} u_{\ell'} u_{\ell+\ell'} = \frac{1}{N\Delta\ell} \sum_k \mathrm{e}^{irk\ell} |\hat{u}(k)|^2. \tag{9.5.19}$$

This expresses the two-point spatial correlation function in terms of the Fourier transform of $|\hat{u}(k)|^2$. It is convenient to consider the continuum limit $\Delta\ell \to 0, N \to \infty$ with $N\Delta\ell = 2\pi$. However, it is first necessary to renormalize the fields so that

$$v(x = \ell) = N u_\ell, \quad \hat{v}(k) = \frac{2\pi}{\Delta\ell}\hat{u}(k).$$

The Fourier transforms become

$$\hat{v}(k) = \frac{2\pi}{N}\sum_x e^{-ikx}v(x), \quad v(x) = \frac{1}{2\pi}\sum_k e^{ikx}\hat{v}(k)$$

with $\Delta k = 1, \Delta x = 2\pi/N$. In the continuum limit we have

$$\hat{v}(k) = \int_{-\pi}^{\pi} e^{-ikx}v(x)dx, \quad v(x) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{ikx}\hat{v}(k)$$

and Eq. (9.5.19) becomes

$$\int_{-\pi}^{\pi} v(y)v(x+y)dy = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{ikx}|\hat{v}(k)|^2.$$

Taking averages of both sides and using Eq. (9.5.18),

$$\int_{-\pi}^{\pi} \langle v(y)v(x+y)\rangle dy = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{ikx}\frac{1}{1+\Gamma_D^2k^2}, \quad \Gamma_D = \sqrt{D\Omega u^*/2K_{\text{off}}}.$$

Finally, estimating the discrete sum using

$$\frac{1}{2\pi}\sum_{k=-\infty}^{\infty} e^{ikx}\frac{1}{1+\Gamma_D^2k^2} \approx \int_{-\infty}^{\infty} e^{ikx}\frac{1}{1+\Gamma_D^2k^2}\frac{dk}{2\pi} = \frac{1}{2\Gamma_D}e^{-|x|/\Gamma_D},$$

we conclude that the averaged correlation function is a decaying exponential with correlation length Γ_D, which establishes that there exist localized states provided that $\Gamma_D \ll 2\pi$ [429].

Box 9E. Fourier analysis of positive feedback model of cell polarization.

The FP equation corresponding to the Langevin equation (9.5.13) is

$$\frac{\partial p}{\partial t} = -\sum_\ell \frac{\partial[V_\ell(\mathbf{u})p(\mathbf{u},t)]}{\partial u_\ell} + \frac{1}{2\Omega}\sum_{\ell,\ell'}\frac{\partial^2[D_{\ell\ell'}(\mathbf{u})p(\mathbf{u},t)]}{\partial u_\ell \partial u_{\ell'}}. \tag{9.5.20}$$

The Fourier transform of the FP equation (9.5.20) with $p(\mathbf{u},t) = P(\hat{\mathbf{u}},t)$, which can be obtained by considering the Fourier transform of the Langevin equation (9.5.13), see Ex. 9.8, is then

$$\frac{\partial P}{\partial t} = -\sum_k \frac{\partial[\hat{V}(k)P(\hat{\mathbf{u}},t)]}{\partial \hat{u}(k)} + \frac{1}{2\Omega}\sum_{k,k'}\frac{\partial^2[\hat{D}(k,k')P(\hat{\mathbf{u}},t)]}{\partial \hat{u}(k)\partial \hat{u}(k')}, \tag{9.5.21}$$

where

$$\hat{V}(k) = \left[K_{\rm fb}(1-(\Delta\ell)^{-1}\hat{u}(0)) - K_{\rm off} + 2\alpha(\cos(k\Delta\ell)-1)\right]\hat{u}(k), \tag{9.5.22}$$

and

$$\hat{D}(k,k') = \left[K_{\rm fb}(1-(\Delta\ell)^{-1}\hat{u}(0)) + K_{\rm off}\right]\hat{u}(k+k'). \tag{9.5.23}$$

In the regime $\alpha \ll K_{\rm fb} - K_{\rm off}$, the eigenvalues of Eq. (9.5.17) satisfy $\lambda_0 \ll \lambda_k < 0$ so that the mode $\hat{u}(0)$ is relatively stable around the steady-state value $\Delta\ell u^*$. Therefore, conditioning the stochastic process by fixing $\hat{u}(0) = \Delta\ell u^*$ leads to an effective FP equation for the nonzero Fourier modes $\hat{u}(k), k \neq 0$ [429]:

$$\frac{\partial P_{\rm eff}(\mathbf{u},t)}{\partial t} = -\sum_{k\neq 0}\frac{\partial[\hat{V}_{\rm eff}(k)P_{\rm eff}(\hat{\mathbf{u}},t)]}{\partial\hat{u}(k)} + \frac{1}{2\Omega}\sum_{k,k'\neq 0}\frac{\partial^2[\hat{D}_{\rm eff}(k,k')P^{\rm eff}(\hat{\mathbf{u}},t)]}{\partial\hat{u}(k)\partial\hat{u}(k')}, \tag{9.5.24}$$

where

$$\hat{V}_{\rm eff}(k) = 2\alpha(\cos(k\Delta\ell)-1)\hat{u}(k), \quad \hat{D}_{\rm eff}(k,k,') = 2K_{\rm off}\left[\Delta\ell\,\hat{u}(k+k') - \frac{\hat{u}(k)\hat{u}(k')}{u^*}\right]. \tag{9.5.25}$$

Multiplying Eq. (9.5.24) by $\hat{u}(k)$ and integrating by parts with respect to $\hat{\mathbf{u}}$ leads to an ODE for the first moment:

$$\frac{d\langle\hat{U}(k,t)\rangle}{dt} \equiv \int \hat{u}(k)P^{\rm eff}(\hat{\mathbf{u}},t)d\hat{\mathbf{u}} = 2\alpha\langle\hat{U}(k,t)\rangle(\cos(k\Delta\ell)-1),$$

which implies that $\langle\hat{U}(k,t)\rangle \to 0$ as $t\to 0$ for all $k\neq 0$. Repeating for the second moments by multiplying Eq. (9.5.24) by $\hat{u}(k)\hat{u}(k')$ and integrating by parts gives (see Ex. 9.8)

$$\frac{d\langle\hat{U}(k,t)\hat{U}(k',t)\rangle}{dt} = \frac{K_{\rm off}\Delta\ell}{\Omega}\langle\hat{U}(k+k',t)\rangle + \left[2\alpha(\cos(k\Delta\ell)+\cos(k'\Delta\ell)-2) - \frac{K_{\rm off}}{\Omega u^*}\right]\langle\hat{U}(k,t)\hat{U}(k',t)\rangle.$$

Since, $\langle\hat{U}(k+k',t)\rangle = 0$ unless $k' = -k$, we see that all two-point correlations $\langle\hat{U}(k,t)\hat{U}(k',t)\rangle \to 0$ as $t\to 0$ unless $k' = -k$, in which case

$$\langle\hat{U}(k,t)\hat{U}(-k,t)\rangle = \langle|\hat{U}(k,t)|^2\rangle \to \frac{(u^*\Delta\ell)^2}{1+4\alpha\Omega(1-\cos(k\Delta\ell))u^*/K_{\rm off}}. \tag{9.5.26}$$

For large N and small $\Delta\ell$, we can take $\cos(k\Delta\ell)\approx 1-(k\Delta\ell)^2/2$ resulting in the Lorentzian distribution (9.5.18).

9.5.3 Active Transport Models

A complementary mechanism for cell polarization in yeast is shown in Fig. 9.22. Here the asymmetric distribution of the signaling molecule Cdc42 within the plasma membrane and the orientation of actin filaments are mutually enhanced through a positive feedback loop resulting in the formation of polarized distribution of Cdc42 and oriented actin cables [415, 597]. In order to develop the basic theory, it is convenient to treat the polarization region as a disk of radius r_0 at the center of a planar membrane so that curvature effects can be ignored. The density $u(\mathbf{x},t)$ of signaling molecules within the plasma membrane is then taken to evolve according to the diffusion equation [415, 597]

$$\frac{\partial u}{\partial t} = D\nabla^2 u - k_- \chi u - \bar{k}_-(1-\chi)u + k_+ \chi U_c, \tag{9.5.27}$$

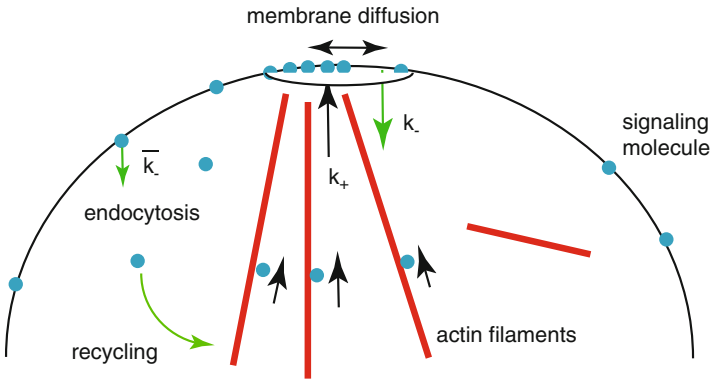


Fig. 9.22: Active transport model of cell polarization. Signaling molecules can attach and orient actin filaments that deliver vesicles carrying the signaling molecule from the cytoplasm to the plasma membrane. The additional signaling molecules orient more actin filaments that transport more molecules in a positive feedback loop, resulting in a polarization region of higher molecule density. The local clustering of actin patches of actin filaments also increases the rate of endocytosis within the polarization domain. Redrawn from [415]

where k_- and $\bar{k}_-$ are the rates of endocytosis inside and outside the polarization domain respectively, k_+ is the rate of delivery to the polarization domain, U_c is the total number of cytosolic molecules, and χ is an index function such that $\chi(\mathbf{x}) = 1$ if $|\mathbf{x}| < r_0$ and zero otherwise. Note that Eq. (9.5.27) describes the dynamics of molecular transport once polarization has been established, rather than during the initial stages of polarization—the main goal is to establish that such a system can maintain a polarized state in steady state. Given the radial symmetry of the polarization

domain, and ignoring any boundaries of the membrane, we can set $u(\mathbf{x},t) = u(r,t)$ with

$$\nabla^2 = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right).$$

Finally, conservation of the total number of signaling molecules implies that

$$U_{\text{total}} = U_c(t) + 2\pi\int_0^\infty u(r,t)rdr. \tag{9.5.28}$$

The steady-state equation ($\partial u/\partial t = 0$) can be solved by partitioning the planar membrane into two regions, $0 \le r \le r_0$ and $r_0 \ge 0$, and matching solutions at the circular boundary $r = r_0$ [597]. Denoting the solution in the first and second regions by $u_1(r)$ and $u_2(r)$, respectively, we have

$$u_1''(r) + \frac{1}{r}u_1'(r) - k_- u_1(r) + k_+ U_c, \quad 0 \le r \le r_0,$$

and

$$u_2''(r) + \frac{1}{r}u_2'(r) - \bar{k}_- u_1(r) = 0, \quad r \ge r_0,$$

together with the boundary conditions

$$u_1'(0) = 0, \quad \lim_{r\to\infty} u_2(r) = 0,$$

and the continuity conditions

$$u_1(r_0) = u_2(r_0) = U, \quad u_1'(r_0) = u_2'(r_0)$$

for some unknown constant U. The solutions of the equations for u_1, u_2 take the form

$$u_1(r) = \frac{k_+U_c}{k_-} + \left(U - \frac{k_+U_c}{k_-}\right)\frac{I_0(\sqrt{k_-}r)}{I_0(\sqrt{k_-}r_0)}, \quad u_2(r) = U\frac{K_0\left(\sqrt{\bar{k}_-}r\right)}{K_0(\sqrt{\bar{k}_-}r_0)}, \tag{9.5.29}$$

where $I_k(r)$ and $K_k(r)$ are the modified Bessel functions of the second kind. The unknown U is determined by matching the first derivatives at the boundary $r = r_0$ and using standard properties of modified Bessel functions:

$$\left(U - \frac{k_+U_c}{k_-}\right)\frac{\sqrt{k_-}I_1(r_0\sqrt{k_-})}{I_0(r_0\sqrt{k_-})} = -U\frac{\sqrt{\bar{k}_-}K_0\left(r_0\sqrt{\bar{k}_-}\right)}{K_0(r_0\sqrt{\bar{k}_-})},$$

which leads to the result

$$U = \frac{k_+U_c}{k_-}\Gamma(k_-,\bar{k}_-),$$

with

$$\Gamma(k_-,\bar{k}_-)=\frac{\sqrt{k_-}I_1(r_0\sqrt{k_-})}{I_0(r_0\sqrt{k_-})}\left[\frac{\sqrt{k_-}I_1(r_0\sqrt{k_-})}{I_0(r_0\sqrt{k_-})}+\frac{\sqrt{\bar{k}_-}K_1(r_0\sqrt{\bar{k}_-})}{K_0(r_0\sqrt{\bar{k}_-})}\right]^{-1}.$$

Finally, the steady-state number of molecules in the cytoplasm can be determined self-consistently using the conservation Eq. (9.5.28). A plot of the solution against r shows a unimodal function with a peak within the polarization domain at $r=0$ [597].

Numerical simulations confirm that a stable spatially localized distribution of signaling molecules within the plasma membrane can be maintained. Moreover, the degree of polarization can be optimized by varying the rates of endocytosis [415]. One limitation of the model, however, is that the packaging of signaling molecules into discrete vesicles is ignored, that is, the model treats transport as a continuous flux of proteins. As highlighted by [374, 560], incorporating vesicular transport into the model makes cell polarization more difficult to sustain. A simple argument for this proceeds as follows. First, it is clear that if the concentration of Cdc42 within a vesicle is the same as a local region of membrane, then fusion of the vesicle releases both Cdc42 and additional lipid membrane so the concentration doesn't change, in contrast to a continuous flux of Cdc42 alone. Hence, exocytic vesicles need to have higher concentrations of the signaling molecule than the polarization site in order to enhance the concentration. A dynamic equilibrium of recycling can only be maintained if endocytic vesicles also have an enhanced concentration of signaling molecules. Although there are various active mechanisms for enhancing the concentration of proteins within vesicles, evidence for such processes within the context of cell polarization is currently lacking.

Another modeling study has investigated conditions under which active transport can generate spontaneous cell polarization [254]. These authors consider a general mechanism of cell polarization based on the positive coupling between membrane-bound signaling models and the cytoskeleton. They demonstrate that the geometry of the organization of cytoskeletal filaments plays a crucial role in determining whether the cell is capable of spontaneous cell polarization or only polarizes in response to an external chemical gradient. More specifically, they show that the former holds if filaments are nucleated at sites on the cell membrane (the actin cytoskeleton), whereas the latter applies if the filaments nucleate from organizing sites within the cytoplasm (microtubule asters). The model thus captures differences in experimental studies of cell polarization in budding yeast [415, 597] and neuron growth cones [59]. The basis model geometry is illustrated in Fig. 9.23. For simplicity, the cell is taken to be two-dimensional and curvature effects are ignored. The cell boundary is given by the x-axis and the cytoplasm given by the half-plane $(x,z), x>0$. Let $u(x,t)$ denote the concentration of signaling molecules in the membrane and let $c(x,z,t)$ denote the corresponding concentration in the cytoplasm. The reaction–diffusion model takes the form

$$\frac{\partial u(x,t)}{\partial t} = D_m \frac{\partial^2 u(x,t)}{\partial x^2} + k_{\text{on}} c(x,0,t) - k_{\text{off}} u(x,t), \tag{9.5.30a}$$

$$\frac{\partial c(x,z,t)}{\partial t} = D\nabla^2 c(x,z,t) - \mathbf{v}\cdot\nabla c(x,z,t). \tag{9.5.30b}$$

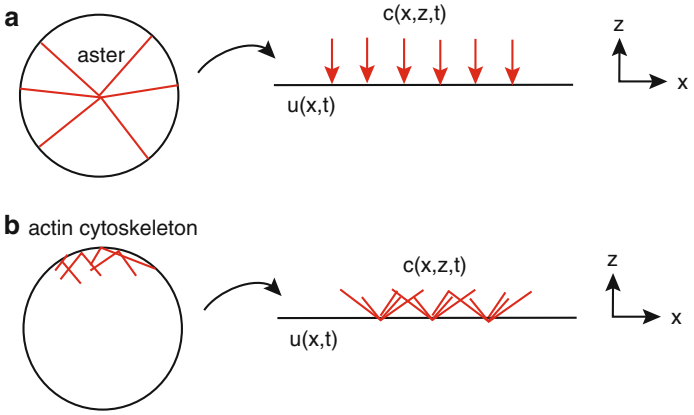


Fig. 9.23: Schematic illustration of filament geometry in model of Hawkins et al. [254]. (**a**) Nucleation at the cell center. (**b**) Nucleation at the cell membrane

The first equation represents diffusion of signaling molecules within the membrane together with transfer between the membrane and cytoplasm, where k_{on} and k_{off} are the binding and unbinding rates. The second equation is an advection–diffusion equation that describes the hybrid transport dynamics of molecules in the cytoplasm, which randomly switch between diffusive motion and ballistic motion along filaments. The advection–diffusion model could be derived from a more detailed model using a QSS approximation (see Sect. 7.4).

The velocity field $\mathbf{v}(x,z,t)$ depends on the geometry of the filaments, which is itself determined by the concentration of signaling molecules on the membrane. Hawkins et al. [254] distinguish between two cases (see also Fig. 9.23):

(a) Filaments that grow from a nucleating center in the cytoplasm (microtubule aster) are approximately perpendicular to the membrane surface. Assuming that the speed of active transport at (x,z) is proportional to the local density of parallel filaments, and that the latter is proportional to the concentration of surface signaling molecules $u(x,t)$, we have

$$\mathbf{v}(x,z,t) = -\alpha u(x,t)\mathbf{e}_z, \tag{9.5.31}$$

where α is a constant that specifies the level of coupling between the signaling molecules and filaments. This type of geometry holds for the distribution of microtubules in neuron growth cones, where GABA receptors appear to associate with and regulate the growing microtubule ends [59].

(b) Filaments that nucleate from sites on the membrane can be approximated by a superposition of asters. Assuming that the velocity field at $\mathbf{r} = (x,z)$ is determined by the local density of filaments, and this decreases with distance from each nucleation site $\mathbf{r}' = (x,0)$, then

$$\mathbf{v}(\mathbf{r},t) = -\alpha \int_{-L/2}^{L/2} \frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|^2} u(x',t)dx', \tag{9.5.32}$$

where L is the "length" of the cell. This geometry reflects the organization of the actin cytoskeleton in budding yeast, as illustrated in Fig. 9.22.

The above equations are supplemented by the conservation equation

$$M = \int_{-L/2}^{L/2} u(x,t)dx + \int_{-L/2}^{L/2}\int_0^\infty c(x,z,t)dz, \tag{9.5.33}$$

with M the total number of signaling molecules. Since the concentration profile decays exponentially in the z direction, the range of z is taken to be the half-line. Finally, we have conservation of flux at the boundary $z = 0$:

$$-D\frac{\partial c(x,0,t)}{\partial z} + v_z(x,0,t)c(x,0,t) + k_{\text{on}}c(x,0,t) - k_{\text{off}}u(x,t) = 0. \tag{9.5.34}$$

The system (9.5.30) has an x-independent steady-state solution $c(x,z,t) = c(z), u(x,t) = u_0$ satisfying the pair of equations

$$0 = k_{\text{on}}c(0) - k_{\text{off}}u_0,$$
$$0 = D\frac{d^2c(z)}{dz^2} - v_z\frac{dc(z)}{dz}.$$

In case (a) we have $v_z = -\alpha u_0$ so that (imposing zero flux at the membrane surface)

$$u_0 = \frac{k_{\text{on}}}{k_{\text{off}}}c(0), \quad c(z) = c(0)\mathrm{e}^{-\xi z}, \quad \xi = \alpha u_0/D.$$

The constant $c_0(0)$ may be determined from the conservation condition. The same solution holds in case (b), except that $\xi = \alpha u_0\pi/D$. In order to show this, assume that L is sufficiently large so that

$$v_z \approx -\alpha u_0 \int_{-\infty}^{\infty} \frac{z}{(x-x')^2+z^2}dx'.$$

Performing the change of variables $x - x' = z\tan\theta$ establishes that the integral has the value π. The stability of the steady state is determined by substituting

$$u(x,t) = u_0 + U(k)\mathrm{e}^{ikx+\lambda t}, \quad c(x,z,t) = c(z) + C(k,z)\mathrm{e}^{ikx+\lambda t}$$

into Eq. (9.5.30) and Taylor expanding to first order in $U(k)$ and $C(k,z)$. The resulting linear equations are

$$(\lambda + D_m k^2 + k_{\text{off}})U(k) = k_{\text{on}} C(k,0), \tag{9.5.35a}$$

$$(\lambda + Dk^2)C(k,z) = D\frac{d^2C(k,z)}{dz^2} - ikv_{0,x}C(k,z) - v_{0,z}\frac{dC(k,z)}{dz} - \beta U(k)\frac{dc_0(z)}{dz}, \tag{9.5.35b}$$

where $\mathbf{v}_0(\mathbf{r},t) = \mathbf{v}(\mathbf{r},t)|_{u=u_0}$ and $(\mathbf{v} - \mathbf{v}_0)\dot{\mathbf{e}}_z \approx \beta U(k)\mathrm{e}^{ikx+\lambda t}$. For the velocity field (9.5.32) it can be shown that Eq. (9.5.35b) has a solution of the form (see Ex. 9.9)

$$C(k,z) = \left[a(k)\mathrm{e}^{-\xi z} + (1-a(k))e^{-\rho z}\right]C(k,0),$$

with

$$\rho = \frac{1}{2}\left[\xi \pm \sqrt{\xi^2 + 4k^2 + 4\lambda/D}\right] \tag{9.5.36}$$

and $a(k)$ a function of k.

Substituting the linearized solutions into the zero-flux condition (9.5.34) then yields the following characteristic equation for λ (see Ex. 9.9):

$$(\lambda - D\xi k)\left(k_{\text{off}}(2\xi - \rho) + (\lambda + D_m k^2)(-k_{\text{on}}/D + \xi - \rho)\right) + (\xi + k - \rho)\xi^2 D k_{\text{off}} = 0 \tag{9.5.37}$$

Numerically solving this equation for $\lambda = \lambda(k)$, one finds that $\lambda(k) < 0$ for all k, which means that the steady state is stable. On the other hand, repeating the analysis for the vector field (9.5.32) leads to the characteristic equation (see Ex. 9.9)

$$(\lambda - D\xi k)\left(k_{\text{off}}(2\xi - \rho) + (\lambda + D_m k^2)(-k_{\text{on}}/D + \xi - \rho)\right) + (\xi + k - \rho)\xi^2 D k_{\text{off}} = 0. \tag{9.5.38}$$

Plotting the corresponding dispersion curve $\lambda = \lambda(k)$ now shows that the steady state is unstable, with $\lambda(k)$ having a positive maximum at a nonzero wavenumber $k_{\max}$ (see Fig. 9.24a). One also finds that k_c increases with ξ and thus both with the strength of coupling α and the total number of signaling molecules M (since u_0 increases with M). For a given cell circumference L, the number of polarization patches will be n when $n \leq k_{\max}L/2\pi < n+1$. This implies that the number of polarization patches grows with ξ, as illustrated in Fig. 9.23b. Therefore, the model provides a possible explanation of why spontaneous polarization occurs in budding yeast, where signaling molecules such as Cdc42 interact with the actin cytoskeleton, but not in neural growth cones, where GABA receptors interact with microtubules. Hawkins et al. [254] suggest that these differences might reflect the contrasting demands on the two cell types during chemical gradient sensing. That

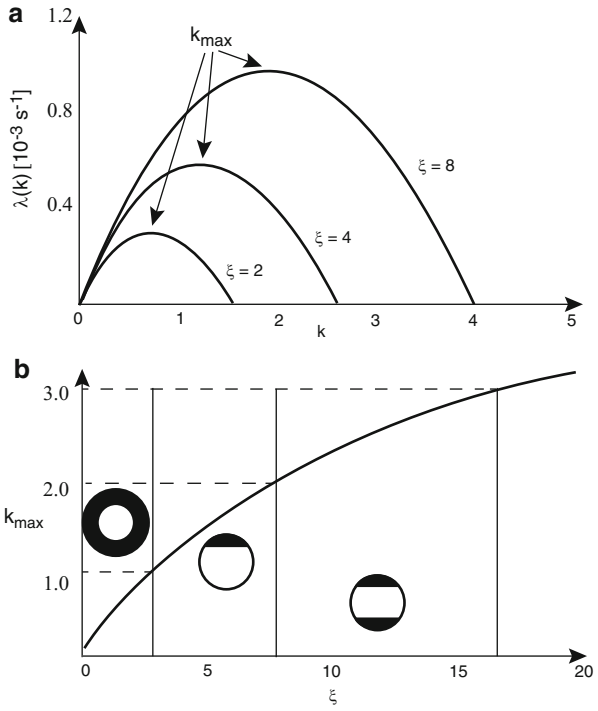


Fig. 9.24: Active transport model of cell polarization. (**a**) Sketch of dispersion curves $\lambda = \lambda(k)$ for the RD system (9.5.30) with aster-like velocity field (9.5.32) and various strengths of coupling (since ξ is proportional to α). Parameter values are $k_{on} = 1\,\mu\text{ms}^{-1}$, $k_{off} = 0.1\,\text{s}^{-1}$, $D = 0.1\,\mu\text{m}^2\text{s}^{-1}$, $D_m = 0.01\,\mu\text{m}^2\text{s}^{-1}$, and $L = 2\pi R$ with cell radius $R = 10\,\mu\text{m}$. Both k and ξ are in units of R^{-1}. (**b**) Sketch of how the dominant wavenumber $k_{\max}$ varies with ξ (in units of R^{-1}). For weak coupling (small ξ), $k_{\max} < 1$ and there are no polarization patches since one cannot fit a single period of a sinusoid within the domain of length L. However, as ξ increases the wavelength of the growing pattern decreases and one can fit n periods into the domain when $n < k_{\max} < n+1$, that is, there are n polarization patches. Redrawn from [254]

is, neuron growth cones respond to relatively weak chemical gradients and thus need to suppress spontaneous fluctuations, whereas budding yeast tends to operate in relatively large gradients that can override any spontaneous fluctuations. This is particularly important for yeast, since they need to polarize quickly in order to gain a mating partner, which means operating in a regime with large coupling α.

9.5.4 Turing-Like Mechanism for Cell Polarization

One mechanism for generating a stable polarization pattern in the absence of noise is to include the effects of an inhibitor, which leads to a reaction–diffusion system that

supports a Turing instability [231, 437] (see Sect. 9.2). For concreteness, we will describe the formulation of Menchon et al. [437], which has been used to model neuronal polarization. The basic idea is to extend the positive feedback model of Altschuler et al. [7, 300] by including a second diffusing molecule which modulates the rate of endocytosis of a membrane-bound molecule such as Cdc42. Let u and v

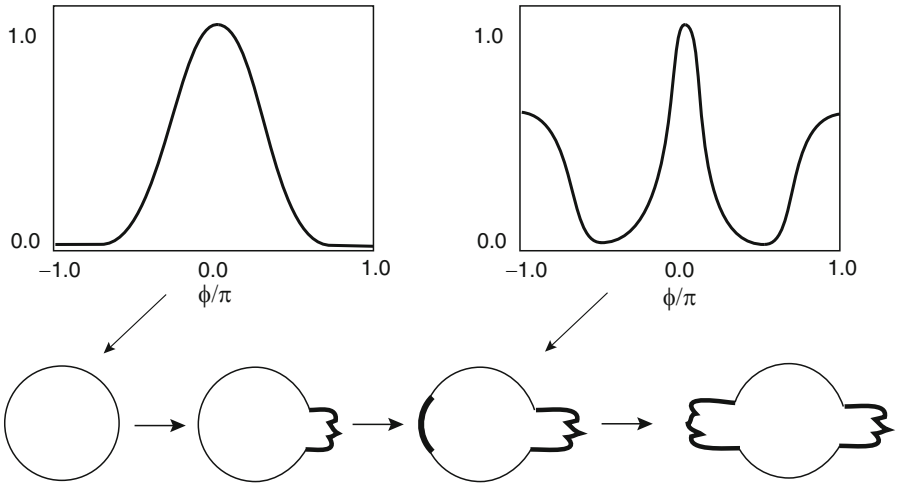


Fig. 9.25: Schematic illustration of neuronal bipolarity

denote the cell membrane concentrations of membrane proteins and modulators of endocytosis, respectively. The corresponding RD system is

$$\frac{\partial u}{\partial t} = D_u \nabla^2 u + k_{\text{on}}(1 + r_1 u) - k_{\text{off}} \frac{uv}{1 + r_2 u}, \tag{9.5.39a}$$

$$\frac{\partial v}{\partial t} = D_v \nabla^2 v + k'_{\text{on}}(1 + r_3 u) - k'_{\text{off}} v. \tag{9.5.39b}$$

Both molecular species are recruited to the membrane by a combination of spontaneous association and positive feedback mediated by the membrane-bound protein. However, the rate of endocytosis of the membrane protein is described by Michaelis-Menten kinetics due to the enzymatic action of the endocytosis modulator. We can identify u as the activator and v as the inhibitor. Consider the simplified case of a two-dimensional cell with a circular membrane of radius R. Introducing polar coordinates (r, ϕ), we can treat the spatial domain as an interval of length $L = 2\pi R$ with periodic boundary conditions. It is also convenient to nondimensionalize by

defining $\tau = tD_u/L^2$ and absorbing a factor of r_1 into u and v. The system (9.5.39) then reduces to

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \gamma f(u,v), \quad f(u,v) = (1+u) - a_1 \frac{uv}{1+b_1 u}, \tag{9.5.40a}$$

$$\frac{\partial v}{\partial t} = D\frac{\partial^2 v}{\partial x^2} + \gamma g(u,v), \quad g(u,v) = a_2(1+u) - a_3 v, \tag{9.5.40b}$$

with $x \in [0,1]$ and

$$D = \frac{D_v}{D_u}, a_1 = \frac{k_{\text{off}}}{k_{\text{on}} r_1^2}, a_3 = \frac{k'_{\text{on}}}{k_{\text{on}}}, a_3 = \frac{k'_{\text{off}}}{k_{\text{on}} r_1}, b_1 = \frac{r_2}{r_1}, b_2 = \frac{r_3}{r_1}, \gamma = \frac{r_1 k_{\text{on}} L^2}{D_u}.$$

The conditions for a Turing instability can now be derived as outlined in Sect. 9.2. Menchon et al. [437] found parameter regimes, in which the dominant spatial pattern had a single peak around the circular membrane, corresponding to a single site of cell polarization. (This is analogous to the growth of the first Fourier mode $e^{\pm i\theta}$ in the ring model considered in Ex. 9.3.) In the case of a developing neuron, such a site would indicate where a single neurite (precursor axon or dendrite) starts to grow. Interestingly, if neurite growth is incorporated into the model of cell polarization, one finds that the concentration of membrane protein sharpens and then spontaneously changes into a double-peaked profile, signaling the formation of a second polarization site at the opposite pole (see Fig. 9.25). This is consistent with the experimental observation that a second neurite tends to form on the opposite pole to the first neurite. Interestingly, it is also possible to induce yeast cells to make two buds [280].

9.5.5 *Wave-Pinning and Cell Polarity in a Bistable Reaction-Diffusion Model*

We end this section by considering a reaction–diffusion model of cell polarity in motile eukaryotic cells developed and analyzed by Keshet et al. [301, 452, 453]; see also [484, 489]. The simplest version of the model considers a single Rho GTPase that can transition between inactive and active forms diffusing in a bounded 1D domain of length L [452]; see Fig. 9.19a. Let $a(x,t)$ and $b(x,t)$ be the concentrations of the active/inactive states. Then

$$\frac{\partial a}{\partial t} = D_a \frac{\partial^2 a}{\partial x^2} + f(a,b), \tag{9.5.41a}$$

$$\frac{\partial b}{\partial t} = D_b \frac{\partial^2 b}{\partial x^2} - f(a,b). \tag{9.5.41b}$$

Since the rate of diffusion of the membrane-bound (active) state is significantly slower than that of the cytosolic (inactive) state, $D_a \ll D_b$. The nonlinear function $f(a,b)$ represents the difference between the rates of activation and inactivation of

the Rho GTPase. Assuming there is cooperative positive feedback in the activation of the protein, which is modeled as a Hill function of index 2, then

$$f(a,b) = b\left(k_0 + \frac{\gamma a^2}{K^2 + a^2}\right) - k_- a. \tag{9.5.42}$$

It can be checked that for a range of uniform concentrations of the inactive state, $b_{\min} < b < b_{\max}$, the space-clamped version of the model exhibits bistability with two stable fixed points $a_\pm(b)$ separated by an unstable fixed point $a_0(b)$. Equation (9.5.41) is supplemented by no-flux boundary conditions at the ends of the domain:

$$\left.\frac{\partial a}{\partial x}\right|_{x=0,L} = 0, \quad \left.\frac{\partial b}{\partial x}\right|_{x=0,L} = 0. \tag{9.5.43}$$

It follows that there is mass conservation of the total amount of Rho GTPase, that is,

$$\int_0^L (a+b)dx = C. \tag{9.5.44}$$

The emergence of cell polarization in this model can be understood in terms of front propagation in a bistable reaction–diffusion system (see Box 9F), with the following additional features [452, 453]: (i) the inactive and active states have unequal rates of diffusion; (ii) the total amount of each GTPase is conserved. Consequently, a local stimulus induces a propagating front that decelerates as it propagates across the cell so that it becomes stationary, a process known as *wave-pinning*; the stationary front persists in the absence of the stimulus and represents a polarized cell. Note that the wave-pinning mechanism is distinct from the diffusion-driven Turing mechanism (Sect. 9.5.4). One of the basic differences is that in the latter class of models, a homogeneous state becomes unstable to arbitrarily small fluctuations and the nonlinear reaction terms support the growth of a spatially varying pattern via a Turing stability. The mathematical explanation of wave-pinning proceeds as follows [452]. First, since $D_b \gg D_a$ and there are no-flux boundary conditions, one can assume that b rapidly diffuses to establish a uniform concentration within the bounded domain $[0,L]$; b then changes on a slower time scale as the a dynamics evolves (QSS approximation). Thus, on short time scales b can be treated as a fixed global parameter of a scalar equation for $a(x,t)$ given by Eq. (9.5.41a). Suppose that initially $b_{\min} < b < b_{\max}$, so Eq. (9.5.41a) is bistable. On an infinite domain, the bistable equation supports the propagation of a traveling front linking the stable fixed point $a_+(b), a_-(b)$ (see Box 9F). That is, for $-\infty < x < \infty$, there exists a monotonically decreasing solution $a(x,t) = A(\xi)$, $\xi = x - ct$ with $\lim_{\xi\to-\infty} A(\xi) = a_+(b)$ and $\lim_{\xi\to\infty} A(\xi) = a_-(b)$. Moreover the wavespeed satisfies $c = c(b)$ with

$$c(b) = \frac{\int_{a_-}^{a_+} f(a,b)da}{\int_{-\infty}^{\infty} (\partial A/\partial\xi)^2 d\xi}. \tag{9.5.45}$$

Note that the wavespeed depends on the global parameter b. Since the denominator of Eq. (9.5.45) is always positive, the sign of $c(b)$ will depend on the sign of $I(b) \equiv \int_{a_-}^{a_+} f(a,b)da$, which has a geometrical interpretation in terms of the difference between the area of the curve $y = f(a,b)$ above the straight line $y = k_- a$ and the area below; see Fig. 9.26. In the case of a sufficiently sharp front that is away from the boundaries, these results carry over to the bounded domain $[0,L]$.

Now suppose that a transient stimulus near the edge of the cell at $x = 0$ triggers at time $t = 0$ a traveling front as described above. This implies that $b_{\min} < b(0) < b_{max}$ and $I(b(0)) > 0$. As the front starts to propagate into the interior of the cell, it converts a greater fraction of the domain from $a \approx a_-(b)$ to $a \approx a_+(b)$. From the conservation condition (9.5.44), it follows that the approximately uniform concentration $b(t)$ of the inactive state decreases, eventually reaching a critical value b_c, $b_{\min} < b_c < b_{\max}$, for which

$$I(b_c) \equiv \int_{a_-}^{a_+} f(a,b_c)da = 0, \tag{9.5.46}$$

and wave-pinning occurs. The basic steps are illustrated in Fig. 9.26. One interesting issue is to what extent the wave-pinning mechanism differs from the phenomenon of wave propagation failure due to spatial discretization. This is particularly important given that any numerical simulation of the wave-pinning model involves the introduction of a spatial grid, and the wave becomes more sensitive to discretization effects as it slows down. A careful numerical study has shown that wave-pinning and propagation failure are distinct effects [667]. In the same study, a stochastic version of the wave-pinning model was also considered, which takes into account fluctuations in the number of active and inactive molecules at low concentrations. It was found that when the total number of molecules is lowered,

wave-pinning behavior is lost due to a broadening of the transition layer as well as increasing fluctuations in the pinning position.

Box 9F. Traveling fronts in a bistable RD equation [66, 322].

Consider a scalar bistable RD equation of the form

$$\frac{\partial a}{\partial t} = \frac{\partial^2 a}{\partial x^2} + f(a), \quad -\infty < x < \infty. \tag{9.5.47}$$

Suppose that $f(a)$ is chosen so that the corresponding ODE, $dv/dt = f(a)$, has stable equilibria at $a = a_\pm$, $a_+ > a_-$, separated by an unstable equilibrium at $a = a_0$. We define a traveling front solution according to

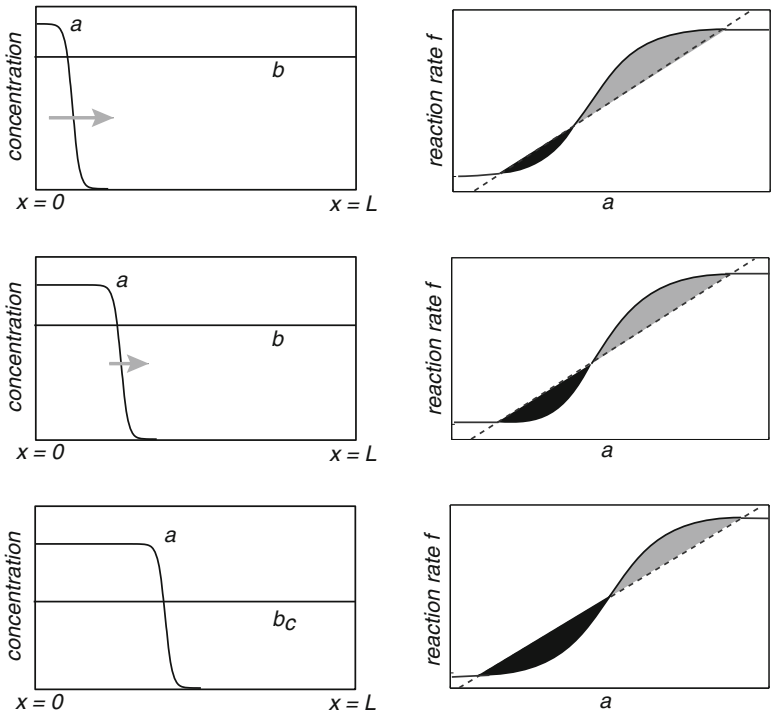


Fig. 9.26: Schematic diagram explaining the mechanism of wave-pinning developed in [452]. A sequence of snapshots of the traveling front (*left column*) showing that as the front advances into the domain, the background concentration b of the inactive state decreases so that the front decelerates until it becomes stationary. The corresponding geometric construction of $I(b)$ (*right column*), which is given by the difference of the *shaded areas*, shows that $I(b)$ is initially positive but vanishes at the critical value b_c

$$a(x,t) = a(x-ct) = A(\xi), \quad \xi = x - ct \tag{9.5.48}$$

for some yet to be determined wavespeed c, supplemented by asymptotic boundary conditions ensuring that the front links the two stable fixed points of the x-independent system. For concreteness, we take

$$A(\xi) \to a_+ \text{ as } \xi \to -\infty, \quad A(\xi) \to a_- \text{ as } \xi \to \infty. \tag{9.5.49}$$

Substituting the traveling front solution into the bistable equation (9.5.47) yields the ODE

$$A_{\xi\xi} + cA_\xi + f(A) = 0, \tag{9.5.50}$$

where $A_\xi = dA/d\xi$. Classical phase-plane analysis can now be used to find a traveling front solution by rewriting the second-order equation in the form

$$A_\xi = Z, \quad Z_\xi = -cZ - f(A). \tag{9.5.51}$$

One now has to look for a solution that links the excited state $(A,Z) = (a_+,0)$ at $\xi \to -\infty$ to the state $a = a_-$ at $\xi \to \infty$—a so-called heteroclinic connection. This can be achieved using a geometric argument based on a shooting method, as illustrated in Fig. 9.27 for the cubic $f(a) = a(a-a_0)(1-a)$ with $a_- = 0, a_+ = 1$ and $0 < a_0 < 1$. Suppose that $0 < a_0 < 1/2$ so that $c > 0$ (see below). First note that irrespective of the speed c, the fixed points $(1,0)$ and $(0,0)$ in the phase plane are saddles, each with one-dimensional stable and unstable manifolds. By looking at trajectories in the phase plane, it is straightforward to see that when $c \ll 1$, the unstable manifold of $(1,0)$ lies below the stable manifold of $(0,0)$ when $0 < A < 1$, whereas the opposite holds when c is very large. Since these manifolds depend continuously on c, it follows that there must exist at least one value of c for which the manifolds cross, and this corresponds to the heteroclinic connection that represents the traveling front solution. It can also be established that this front is unique.

A useful formula for determining the sign of the wave speed can be obtained by multiplying both sides of Eq. (9.5.50) by A_ξ and integrating with respect to ξ:

$$\begin{aligned} c\int_{-\infty}^{\infty}(A_\xi)^2 d\xi &= -\int_{-\infty}^{\infty} A_\xi f(A(\xi))d\xi - \int_{-\infty}^{\infty} A_\xi A_{\xi\xi} d\xi, \\ &= \int_{a_-}^{a_+} f(A)dA, \end{aligned} \tag{9.5.52}$$

since $A(\xi)$ is monotone, and $\int_{-\infty}^{\infty} A_\xi A_{\xi\xi} d\xi = \int_{-\infty}^{\infty} \frac{d[A_\xi^2/2]}{d\xi} d\xi = 0$. As the integral on the left-hand side is positive, it follows that the sign of c is determined by the sign of the area of f between the two stable equilibria. In the case of the cubic, if $0 < a_0 < 1/2$, then the latter is positive and the wave moves to the right. If the negative and positive areas exactly cancel, then the front is stationary (pinned).

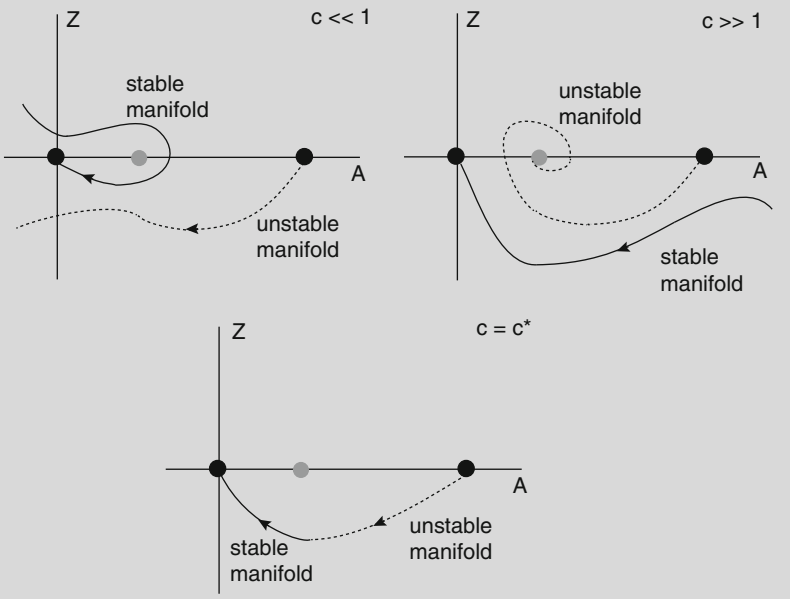


Fig. 9.27: Shooting method for constructing a front solution in the (A, Z) phase plane with $Z = A_\xi$. See text for details

9.6 Exercises

Problem 9.1 (Robustness of a dual protein gradient). The concentration of Pom1 in the membrane of fission yeast evolves according to the equation

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} - \mu c,$$

where μ is the rate of membrane dissociation, supplemented by the boundary conditions

$$D\frac{\partial c}{\partial x}\bigg|_{x=\pm L/2} = \pm J.$$

(a) Show that the steady-state solution is

$$c(x) = \frac{J}{D}\frac{\lambda \cosh(x/\lambda)}{\sinh(L/2\lambda)},$$

where $\lambda = \sqrt{D/\mu}$ is the characteristic decay length of the spatial gradient.

(b) The uncertainty Δx in the position of the center is given by

$$\Delta c = \frac{1}{2}|c''(0)|\Delta x^2. \tag{9.6.53}$$

where $\Delta c = \sqrt{c(0)}/a^2$ and a is the size of the measuring region. Using part (a), show that

$$\Delta x = \left(\frac{4D\lambda^3 \sinh(L/2\lambda)}{Ja^2}\right)^{1/4}. \tag{9.6.54}$$

Problem 9.2 (A little group theory).

(a) Show that the set of rotation matrices

$$\mathbf{M}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad \theta \in [0, 2\pi),$$

and the reflection matrix

$$\mathbf{M}_\kappa = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

together with the rules of matrix multiplication form the representation of a group acting on the linear vector space $\mathbb{R}^2$. The abstract group is $O(2)$. Setting $z = x + iy$, show that rotation by θ corresponds to the transformation $z \to \mathrm{e}^{i\theta} z$ and reflection becomes $z \to z^*$.

(b) Suppose that there exists a matrix representation of some group. Show that a corresponding one-dimensional representation can be obtained by taking the determinants of the matrices. Calculate the representation explicitly for the dihedral group D_3.

(c) By considering a square centered at the origin of the (x,y)-plane construct the matrix representation of the dihedral group D_4, which consists of operations that transform the square into itself. Is the group representation irreducible? Calculate the axial isotropy subgroups of the given group representation.

Problem 9.3 (Pattern formation on a ring). Consider an RD system defined on a ring:

$$\begin{aligned} \frac{\partial u_1}{\partial t} &= D\frac{\partial^2 u_1}{\partial \theta^2} + \kappa(u_1 + a u_2 - u_1 u_2^2 - C u_1 u_2) \\ \frac{\partial u_2}{\partial t} &= \frac{\partial^2 u_2}{\partial \theta^2} + \kappa(-u_1 + b u_2 + u_1 u_2^2 + C u_1 u_2), \end{aligned}$$

with $\theta \in [0, 2\pi]$ and periodic concentrations $u_i(\theta + 2m\pi, t) = u_i(\theta, t)$ for all integers m. Suppose that the onset of a Turing instability of the stationary state at $(u_1, u_2) = (0,0)$ occurs at a critical parameter value $a = a_c$ due to the particular Fourier modes $\mathrm{e}^{\pm in\theta}$ becoming marginally stable. This means that there exists a vector $\mathbf{c}$ such that

$$\mathbf{A}(n)\mathbf{c} \equiv \begin{pmatrix} \kappa - Dn^2 & \kappa a_c \\ -\kappa & b\kappa - n^2 \end{pmatrix} \mathbf{c} = 0.$$

Suppose that the system is just beyond the bifurcation point and set $a - a_c = \varepsilon \Delta a$ with $0 < \varepsilon \ll 1$. Introduce the slow time scale $\tau = \varepsilon t$ with $\partial/\partial t \to \varepsilon \partial/\partial \tau$.

(a) Substitute the perturbation expansion perturbation expansion

$$u_i = \varepsilon^{1/2} u_i^{(1)} + \varepsilon u_i^{(2)} + \varepsilon^{3/2} u_i^{(3)} + \ldots$$

into the RD equation and separately collect terms in powers of $\varepsilon^{1/2}, \varepsilon$, and $\varepsilon^{3/2}$, respectively. Hence, derive the linear inhomogeneous equations

$$\mathbb{L}\mathbf{u}^{(1)} = 0, \quad \mathbb{L}\mathbf{u}^{(2)} = \mathbf{h}^{(2)}, \quad \mathbb{L}\mathbf{u}^{(3)} = \mathbf{h}^{(3)},$$

where

$$\mathbb{L}\mathbf{u} = \begin{pmatrix} \kappa & \kappa a_c \\ -\kappa & b\kappa \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \begin{pmatrix} D & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \partial^2 u_1/\partial \theta^2 \\ \partial^2 u_2/\partial \theta^2 \end{pmatrix}.$$

Determine the vectors $\mathbf{h}^{(2)}$ and $\mathbf{h}^{(3)}$ as functions of $\mathbf{u}^{(1)}$ and $\mathbf{u}^{(2)}$.

(b) Write the $O(\varepsilon^{1/2})$ solution as

$$\mathbf{u}^{(1)}(\theta, \tau) = \left[z(\tau)\, \mathrm{e}^{in\theta} + z^*(\tau)\, \mathrm{e}^{-in\theta} \right] \mathbf{c}.$$

A dynamical equation for the complex amplitude $z(\tau)$ can be obtained by deriving solvability conditions for the higher-order equations. That is, the linear operator $\mathbb{L}$ is self-adjoint with respect to the inner product

$$\langle \mathbf{u} | \mathbf{v} \rangle = \sum_{j=1,2} \int_0^{2\pi} u_j^*(\theta) v_j(\theta) \frac{d\theta}{2\pi}.$$

Since $\mathbb{L}\mathrm{e}^{\pm in\theta}\mathbf{c} = 0$, it follows from the Fredholm alternative theorem (Box 9B) that the higher-order equations only have a solution if the following solvability conditions are satisfied:

$$\langle \hat{\mathbf{c}} \mathrm{e}^{in\theta} | \mathbf{h}^{(m)} \rangle = 0, \quad m = 2, 3$$

where $\hat{\mathbf{c}}^T \mathbf{A}(n) = 0$. Using the identity

$$\int_0^{2\pi} \mathrm{e}^{im\theta} \mathrm{e}^{im'\theta} \frac{d\theta}{2\pi} = \delta_{m+m',0},$$

and the explicit expressions for $\mathbf{h}^{(2,3)}$, show that (i) the solvability condition for $\mathbf{h}^{(2)}$ is automatically satisfied and (ii) the $O(\varepsilon^{3/2})$ solvability condition is

$$\hat{\mathbf{c}} \cdot \langle \mathrm{e}^{in\theta} | \frac{\partial \mathbf{u}^{(1)}}{\partial \tau} \rangle - \kappa \Delta a \hat{\mathbf{c}} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \langle \mathrm{e}^{in\theta} | u_2^{(1)} \rangle$$
$$= \kappa \hat{\mathbf{c}} \cdot \begin{pmatrix} -1 \\ 1 \end{pmatrix} \langle \mathrm{e}^{in\theta} | \left[u_1^{(1)} u_2^{(1)} u_2^{(1)} + C u_1^{(2)} u_2^{(1)} + C u_1^{(1)} u_2^{(2)} \right] \rangle.$$

(c) Show that $\mathbf{u}^{(2)}$ has the general form

$$\mathbf{u}^{(2)}(\theta) = \mathbf{c}^{+} \mathrm{e}^{2ni\theta} + \mathbf{c}^{-} \mathrm{e}^{-2ni\theta} + \mathbf{c}^{0} + \zeta \mathbf{u}^{(1)}(\theta)$$

Substituting for $\mathbf{u}^{(2)}$ into the equation $\mathbb{L}\mathbf{u}^{(2)} = \mathbf{h}^{(2)}$, derive the equations

$$\mathbf{A}(2n)\mathbf{c}^{+} = \kappa C c_1 c_2 z^2 \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \mathbf{A}(2n)\mathbf{c}^{-} = \kappa C c_1 c_2 z^{*2} \begin{pmatrix} 1 \\ -1 \end{pmatrix},$$
$$\mathbf{A}(0)\mathbf{c}^{0} = 2\kappa C c_1 c_2 |z|^2 \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

(The term involving ζ does not contribute to the cubic amplitude equation.)

(d) Combining the results of parts (b) and (c), evaluate the various inner products to obtain the cubic amplitude equation

$$\frac{dz}{d\tau} = z(\tau)(\eta \Delta a - \Lambda |z(\tau)|^2),$$

after absorbing a factor of κ into τ, with

$$\eta = \hat{c}_1 c_2, \quad \Lambda = (\hat{c}_1 - \hat{c}_2) c_1 c_2 \left[3c_2 + \kappa C^2 (c_2 V_1 + c_1 V_2) \right].$$

and

$$\mathbf{V} = \mathbf{A}(2n)^{-1} \begin{pmatrix} 1 \\ -1 \end{pmatrix} + 2\mathbf{A}(0)^{-1} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Problem 9.4 (Pattern formation and $\mathbf{O}(2)$ symmetry). The RD equations of Ex. 9.3 (and their linearization) are equivariant with respect to the action of the group $\mathbf{O}(2)$ of rotations and reflections on the circle S^1. That is, if $\mathbf{u}(\theta, t)$ is a solution, then so is $\mathbf{u}(\theta + \phi, t)$ (rotations) and $\mathbf{u}(-\theta, t)$ (reflections). Applying these transformations to the solution

$$\mathbf{u}(\theta, \tau) = \left[z(\tau)\, \mathrm{e}^{in\theta} + z^*(\tau)\, \mathrm{e}^{-in\theta} \right] \mathbf{c},$$

determine the action of $\mathbf{O}(2)$ on the amplitudes z, z^*. Show that the cubic amplitude equation

$$\frac{dz}{d\tau} = z(\tau)(\eta \Delta a - \Lambda |z(\tau)|^2),$$

is equivariant with respect to this action of $\mathbf{O}(2)$, and determine the z-dependence of the next order term in the amplitude equation.

Problem 9.5 (Discrete Fourier transforms).

(a) Consider the discrete Fourier transform pair

$$V(k) = \sum_{\ell} \mathrm{e}^{-ik\ell} v_\ell, \quad v_\ell = \frac{1}{N} \sum_{k} \mathrm{e}^{ik\ell} V(k)$$

with $\ell = 1, \ldots, N$ and $k = 2\pi m/N, m = 0, \ldots, N-1$. Show that

$$\sum_{\ell=1}^{N} \mathrm{e}^{i(k-k')\ell} = N\delta_{k,k'}$$

for all $k = 2\pi m/N$. Hence, given the definition of $V(k)$, verify the formula for the inverse discrete transform.

(b) Use discrete Fourier transforms to solve the difference equation

$$\hat{\Delta} u_j \equiv u_{j+1} - 2u_j + u_{j-1} = f_j, \quad j = 1, \ldots, N-1,$$

$u_0 = u_N = 0$, in terms of the Fourier coefficients of the discrete function $f_j, j = 1, \ldots, N-1$.

Problem 9.6 (Aggregation model of Min protein oscillations). Consider the simplified model of Min oscillations given by [see Eq. (9.4.6)]

$$\begin{aligned} \frac{\partial \rho_d}{\partial t} &= -\bar{\sigma}_E \rho_d + \bar{k}_D(\rho_{\max} - \rho_{de} - \rho_d) - \frac{\partial J_d}{\partial x} \\ \frac{\partial \rho_{de}}{\partial t} &= -\sigma_d \rho_{de} + \bar{\sigma}_E \rho_d, \end{aligned}$$

where ρ_d and ρ_{de} are the concentrations of membrane-bound MinD and MinD·MinE complexes and J_d is the aggregation current

$$J_d = -D_d \partial \rho_d + \rho_d(\rho_{\max} - \rho_{de} - \rho_d)\left[k_1 \partial_x \rho_d + k_2 \partial_x^3 \rho_d\right].$$

Also $\bar{\sigma}_E = \sigma_E \bar{\rho}_E$ and $\bar{k}_D = k_D \bar{\rho}_D$ with $(\bar{\rho}_D, \bar{\rho}_E)$ the unique fixed point of equations (9.4.5).

(a) Linearize the RD equations about the homogeneous fixed point by setting

$$\rho_d - \bar{\rho}_d = u(q)\mathrm{e}^{\lambda t + iqx}, \quad \rho_{de} - \bar{\rho}_{de} = v(q)\mathrm{e}^{\lambda t + iqx}$$

and Taylor expanding to first order in u, v. Thus obtain the eigenvalue equation

$$\mathbf{A}(q) \begin{pmatrix} u \\ v \end{pmatrix} = \lambda(q) \begin{pmatrix} u \\ v \end{pmatrix},$$

with

$$\mathbf{A}(q) = \begin{pmatrix} -\bar{\sigma}_E - \bar{k}_D - f(q^2) & -\bar{k}_D \\ \bar{\sigma}_E & -\sigma_d \end{pmatrix}$$

and

$$f(q^2) = D_d q^2 - \bar{\rho}_d(\rho_{\max} - \bar{\rho}_{de} - \bar{\rho}_d)(k_1 q^2 - k_2 q^4).$$

(b) Show that the homogeneous fixed point is stable unless k_1 exceeds a critical value $k_{1,c}$. Determine $k_{1,c}$ by finding the smallest value of k_1 for which $\mathrm{Re}\,\lambda(q) = 0$ and show that this occurs at the critical wavenumber

$$q_c^4 = \frac{k_D\bar{\rho}_D + \sigma_d + \sigma_E\bar{\rho}_E}{\bar{\rho}_d(\rho_{\max} - \bar{\rho}_d - \bar{\rho}_{de})k_2}.$$

Hence, show that the critical oscillation frequency $\Omega_c = \mathrm{Im}\,\lambda(q_c)$ is given by

$$\Omega_c^2 = \sigma_D\sigma_E\bar{\rho}_D\bar{\rho}_E - \sigma_d^2.$$

Problem 9.7 (Positive feedback model of cell polarization I). Consider the birth–death process arising from a model of cell polarization, which takes the form (9.5.7) with transition rates

$$\omega_+(n) = (N-n)K_{\rm off}, \quad \omega_-(n) = (N-n)nK_{\rm fb} + nK_{\rm on}.$$

(a) Using the fact that the steady-state solution is given by

$$P_s(n) = P_s(0)\prod_{m=1}^{n}\frac{\omega_+(m-1)}{\omega_-(m)},$$

derive the solution

$$\frac{P_s(n)}{P_s(0)} = \frac{N}{N-n}\frac{N_c^n}{n!}\prod_{m=1}^{n}\left[1+\frac{\gamma}{N-m}\right]^{-1}, \quad N_c = \frac{K_{\rm off}}{K_{\rm fb}}, \quad \gamma = \frac{K_{\rm on}}{K_{\rm fb}}$$

for $n < N$ and

$$P_s(N) = \frac{N_c}{N\gamma}P_s(N-1).$$

(b) Suppose that $\gamma \ll 1$. Show that

$$\gamma \ln\frac{N}{N-n} = \gamma\ln\prod_{k=1}^{n}\frac{N-k+1}{N-k} = \gamma\sum_{k=1}^{n}\frac{1}{N-k} + O(\gamma^2).$$

Hence, obtain the approximations

$$\frac{P_s(n)}{P_s(0)} \approx \left(\frac{N}{N-n}\right)^{1-\gamma}\frac{N_c^n}{n!}, \quad \frac{P_s(N)}{P_s(0)} = \frac{1}{\gamma}\frac{N_c^n}{n!}N^{1-\gamma}.$$

Problem 9.8 (Positive feedback model of cell polarization II). Consider the Langevin equation for the positive feedback model of cell polarization [see also Eq. (9.5.20)]:

$$\frac{dU_\ell(t)}{dt} = V_\ell(\mathbf{U})U_\ell(t) + \Omega^{-1/2}\eta_\ell(t),$$

with white noise terms satisfying

$$\langle \eta_\ell \rangle = 0, \quad \langle \eta_\ell \eta_{\ell'} \rangle = D_{\ell\ell'},$$

such that

$$V_\ell(\mathbf{u}) = K_{\text{fb}} u_\ell (1 - \sum_\ell u_\ell) - K_{\text{off}} u_\ell + \alpha \Delta u_\ell$$

and

$$D_{\ell\ell'}(\mathbf{u}) = \left[K_{\text{fb}} u_\ell (1 - \sum_\ell u_\ell) + K_{\text{off}} u_\ell + O(\alpha) \right] \delta_{\ell,\ell'}.$$

For simplicity, consider a 1D lattice with periodic boundary conditions, $U_{\ell+N\Delta\ell} = U_\ell$ for $\ell = 1, \ldots, N$ and lattice spacing $\Delta\ell$.

(a) Show that the discrete Fourier transform of the Langevin equation is

$$\frac{dU(k,t)}{dt} = V(k,t)U(k,t) + \Omega^{-1/2}\eta(k,t)$$

with

$$\langle \eta(k,t) \rangle = 0, \quad \langle \eta(k,t)\eta(k',t') \rangle = \hat{D}(k,k',t)\delta(t-t'),$$

and determine $V(k,t)$ and $D(k,k',t)$ as functions of $U(k,t)$ and $U(k',t)$. Hence, show that the discrete Fourier transform of the FP equation (9.5.20) is given by Eq. (9.5.21).

(b) Taking first and second moments of the effective FP equation (9.5.24), derive the moment equations

$$\frac{d\langle \hat{U}(k,t) \rangle}{dt} = 2\alpha \langle \hat{U}(k,t) \rangle (\cos(k\Delta\ell) - 1)$$

and

$$\frac{d\langle \hat{U}(k,t)\hat{U}(k',t) \rangle}{dt} = \frac{K_{\text{off}}\Delta\ell}{\Omega} \langle \hat{U}(k+k',t) \rangle + \left[2\alpha(\cos(k\Delta\ell) + \cos(k'\Delta\ell) - 2) - \frac{K_{\text{off}}}{\Omega u^*} \right] \langle \hat{U}(k,t)\hat{U}(k',t) \rangle.$$

Hence, determine the large t behavior of the first and second moments.

Problem 9.9 (Active transport model of cell polarization). Consider the following RD model of cell polarization:

$$\frac{\partial u(x,t)}{\partial t} = D_m \frac{\partial^2 u(x,t)}{\partial x^2} + k_{\text{on}} c(x,0,t) - k_{\text{off}} u(x,t),$$
$$\frac{\partial c(x,z,t)}{\partial t} = D\nabla^2 c(x,z,t) - \mathbf{v}\cdot\nabla c(x,z,t)$$

with $0 \le z < \infty$. The velocity field $\mathbf{v}(x,z,t)$ is taken to be of the form

$$\mathbf{v}(x,z,t) = -\alpha u(x,t)\mathbf{e}_z$$

where α is a constant that specifies the level of coupling between the signaling molecules and filaments. The above equations are supplemented by the conservation equation

$$M = \int_{-L/2}^{L/2} u(x,t)dx + \int_{-L/2}^{L/2}\int_0^{\infty} c(x,z,t)dz,$$

with M the total number of signaling molecules. The steady-state solution is

$$u_0 = \frac{k_{\text{on}}}{k_{\text{off}}} c_0(0), \quad c_0(z) = c_0(0)\mathrm{e}^{-\xi z}, \quad \xi = \alpha u_0/D.$$

(a) Linearize about the steady-state solution by substituting

$$u(x,t) = u_0 + U(k)\mathrm{e}^{ikx+\lambda t}, \quad c(x,z,t) = c_0(z) + C(k,z)\mathrm{e}^{ikx+\lambda t},$$

into the RD equations and Taylor expanding to first order in $U(k)$ and $C(k,z)$. After solving the resulting linear differential equation for $C(k,z)$ show that

$$U(k) = \frac{k_{\text{on}}}{(\lambda + D_m k^2 + k_{\text{off}})} C(k,0),$$
$$C(k,z) = C(k,0)\left[a(k)\mathrm{e}^{-\xi z} + (1-a(k))\mathrm{e}^{-\rho z}\right]$$

where

$$\rho = \frac{1}{2}\left[\xi \pm \sqrt{\xi^2 + 4k^2 + 4\lambda/D}\right],$$

and determine the coefficient $a(k)$.

(b) Linearizing the zero-flux condition

$$-D\frac{\partial c(x,0,t)}{\partial z} - \alpha u(x,t)c(x,0,t) + k_{\text{on}} c(x,0,t) - k_{\text{off}} u(x,t) = 0$$

and using the solutions for $U(k)$ and A in part (b), show that $\lambda(k)$ satisfies the dispersion equation

$$(\lambda + Dk^2)\left(k_{\text{off}}(2\xi - \rho) + (\lambda + D_m k^2)(-k_{\text{on}}/D + \xi - \rho)\right) + (\xi - \rho)\xi^2 D k_{\text{off}} = 0.$$

(c) Repeat the analysis for the alternative velocity field

$$\mathbf{v}(\mathbf{r},t) = -\alpha \int_{-\infty}^{\infty} \frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|^2} u(x',t)dx'.$$

In particular, derive the dispersion relation

$$(\lambda - D\xi k)\left(k_{\text{off}}(2\xi-\rho) + (\lambda + D_m k^2)(-k_{\text{on}}/D + \xi - \rho)\right) + (\xi + k - \rho)\xi^2 D k_{\text{off}} = 0.$$

Chapter 10
The WKB Method, Path-Integrals, and Large Deviations

In Sects. 3.4 and 3.5 we highlighted a limitation of the diffusion approximation of jump Markov processes for large system size N, namely, that it can lead to exponentially large errors in solutions to FPT problems. For large but finite N, we expect the stochastic system to "stay close" to the dynamics of the underlying mass-action kinetics and to converge in the zero-noise limit $\varepsilon \equiv N^{-1} \to 0$ to the deterministic system. (Recall that for large N the random variables of the jump Markov process can be treated as continuous rather than discrete.) In terms of the probability distribution over random trajectories or paths realized by solutions to the master equation, it should concentrate around the deterministic path of the kinetic equations in the limit $\varepsilon \to 0$. The source of the error in the diffusion approximation is that it gives a poor estimate of the probability density as $\varepsilon \to 0$. Such a density has the form of a so-called large deviation principle [195, 574, 640]. In order to give a heuristic definition of the latter, consider some random dynamical system in $\mathbb{R}^n$ for which there exists a well-define probability density functional $P_\varepsilon[x]$ over the different sample trajectories $\{x(t)\}_0^T$ in a given time interval $[0,T]$. A large deviation principle for the random paths is that

$$P_\varepsilon[x] \sim \mathrm{e}^{-S[x]/\varepsilon}, \quad \varepsilon \to 0,$$

where $S[x]$ is known as an action functional. Solving the FPT problem for escape from a fixed point attractor of the underlying deterministic system involves finding the most probable paths of escape, which minimize the action functional with respect to the set of all trajectories emanating from the fixed point (under certain additional constraints). Evaluating the action functional along a most probable path from the fixed point to another point x generates a corresponding quasi-potential $\Phi(x)$. The diffusion approximation generates an inaccurate action functional and associated quasi-potential and thus leads to exponentially large errors in the MFPT. Consider for example a single-variable stochastic process that exhibits bistability in the deterministic limit, that is, there exists a pair of stable fixed points at $x = x_\pm$ separated by an unstable fixed point at $x = x_0$. Given a quasi-potential Φ, the MFPT

P.C. Bressloff, *Stochastic Processes in Cell Biology*, Interdisciplinary Applied Mathematics 41, DOI 10.1007/978-3-319-08488-6_10

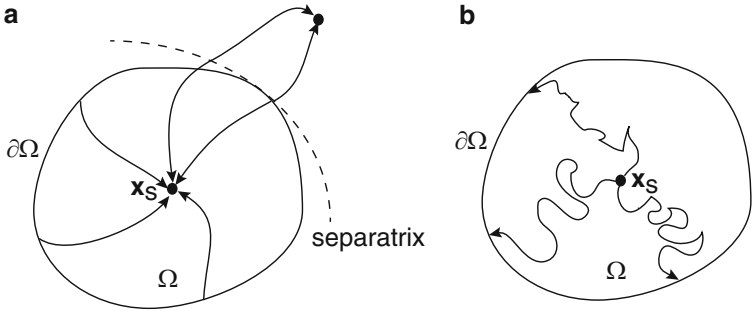


Fig. 10.1: (**a**) Deterministic trajectories of a multistable dynamical system. The subset Ω is contained within the basin of attraction of a fixed point $\mathbf{x}_s$. The boundary of the basin of attraction consists of separatrices, which are also solution trajectories. Paths that start in a neighborhood of a separatrix are attracted by different fixed points, depending whether they begin on the left or right of the separatrix. (**b**) Random trajectories of the stochastic system. Escape from the domain Ω occurs when a random trajectory hits the boundary $\partial\Omega$

τ to escape from the fixed point at x_- takes the general Arrhenius form (see for example (3.3.36))

$$\tau \sim \frac{\Gamma(x_0, x_-)}{\sqrt{|\Phi''(x_0)|\Phi''(x_-)}} \mathrm{e}^{N[\Phi(x_0)-\Phi(x_-)]}, \tag{10.0.1}$$

where Γ is an appropriate prefactor. Moreover, $\Phi(x_0) - \Phi(x_-)$ is the value of the action along the optimal path from x_- to x_0. Thus any errors in the form of the quasi-potential can generate exponentially large errors in the MFPT to escape from a metastable state.

The physical interpretation of the least-action trajectories becomes crucial when solving escape problems in higher dimensions, since a metastable state is now surrounded by a nontrivial boundary (rather than a single point) and one needs to determine the relative weighting of optimal paths crossing different points on the boundary. This is illustrated in Fig. 10.1 for a two-dimensional SDE, whose underlying deterministic system has a stable fixed point $\mathbf{x}_s$ with some basin of attraction $\mathscr{A}$. (A more complex example was encountered in Fig. 10.4.) If $\Omega \subset \mathscr{A}$, then one can extend the MFPT calculation of the scalar case provided that $\partial\Omega$ does not overlap any separatrices [409]; otherwise the analysis is more involved. For a detailed discussion of noise-induced escape problems see the book by Schuss [574].

One efficient method for deriving the correct quasi-potential is to use a WKB approximation of the QSS probability density. This method has been applied to FP equations in the weak noise limit [409, 426, 466, 574], to master equations [145, 152, 163, 252, 341, 653], and more recently to stochastic hybrid systems [72, 74, 321, 470, 475, 476]. In many cases, one can interpret the WKB equation for the quasi-potential in terms of a Hamilton–Jacobi equation, whose corresponding Hamiltonian H is related to the action of large deviation theory according to

$$S[x] = \int_{t_0}^{t} [p\dot{x} - H(x,p)]dt.$$

Thus large deviation theory provides a rigorous foundation for the application and interpretation of WKB methods, in particular, ensuring that the solutions to Hamilton's equations correspond to optimal paths of the full stochastic system.

We begin this chapter by describing how the WKB method can be used to study escape problems in chemical master equations and stochastic hybrid systems (Sect. 10.1). We illustrate the theory by revisiting the example of SAPs, which was analyzed using a linear noise approximation in Sect. 3.5. In subsequent sections we derive action principles underlying the WKB quasi-potentials. However, we avoid the technical aspects of large deviation theory by using path-integrals. We start by considering a path-integral representation of a SDE in the weak noise limit (Sect. 10.2). We then show how to construct the so-called Doi–Peliti path integral representation of a birth–death master equation (Sect. 10.3) and conclude by developing a path integral representation of a stochastic hybrid system (Sect. 10.4). The latter is used to explore the effects of potassium ion channel fluctuations on SAPs in a stochastic Morris–Lecar model (Sect. 10.5).

10.1 The Wentzel-Kramers-Brillouin Method

10.1.1 WKB Method for a Birth–Death Master Equation

Consider the birth–death master equation (3.2.2),

$$\frac{d}{dt}P(n,t) = \omega_+(n-1)P(n-1,t) + \omega_-(n+1)P(n+1,t) - [\omega_+(n) + \omega_-(n)]P(n,t),$$

with $\omega_\pm = N\Omega_\pm$ and reflecting boundary conditions at $n = 0, N$. Suppose that in the deterministic limit $N \to \infty$, the rate equation (3.2.5), $\dot{x} = \Omega_+(x) - \Omega_-(x)$, exhibits bistability, with stable fixed points at $x = x_\pm$ and an unstable fixed point at $x = x_0$. (Throughout the analysis we will switch between n/N and x, with x treated as a continuous variable for large N.) Since we are interested in calculating the MFPT to escape from x_-, for example we impose an absorbing boundary condition at $n = n_0$, that is, $p(n_0,t) = 0$ and take $0 \leq n \leq n_0$. It is convenient to rewrite the master equation (3.2.2) for $n = 0, \ldots, n_0$ as the linear system

$$\frac{d\mathbf{p}}{dt} = \mathbf{A}\mathbf{p} \tag{10.1.2}$$

where $\mathbf{p} = (p_0(t), p_1(t), \ldots, p_{n_0}(t))^T$ with $p_n(t) = p(n,t)$ and $\mathbf{A}$ the transition matrix. Suppose that the matrix $\mathbf{A}$ is irreducible, that is, any state can be reached from any other state through a sequence of transitions. If the absorbing boundary condition at $n = n_0$ were replaced by a reflecting boundary condition, then $\mathbf{A}$ would

have a simple zero eigenvalue with corresponding left eigenvector $\mathbf{1}$ whose components are all unity, that is, $\sum_n A_{nm} = 0$ for all m. The latter follows immediately from conservation of probability in the case of reflecting boundaries. The Perron–Frobenius theorem[1] then ensures that all other eigenvalues are negative and that equation (10.1.2) has a globally attracting steady state ρ_n such that $\sum_m A_{nm}\rho_m = 0$ and $p_n(t) \to \rho_n$ as $t \to \infty$. On the other hand, in the case of an absorbing boundary, probability is no longer conserved since there is an exponentially small but nonzero flux at $n = n_0$ (for large N). The eigenvalues of the transition matrix can now be ordered according to $0 > \lambda_0 \geq \mathrm{Re}[\lambda_1] \geq \mathrm{Re}[\lambda_2] \geq \ldots$ with $|\lambda_0| \sim \mathrm{e}^{-\eta N}$ for $\eta = \mathscr{O}(1)$, whereas λ_r for $r > 0$ are only weakly dependent on N. The exponentially small principal eigenvalue reflects the fact that the flux through the absorbing boundary is exponentially small and in the limit $N \to 0$ reduces to the Perron eigenvalue.

Consider the eigenfunction expansion

$$p_n(t) = \sum_{r=0}^{n_0} C_r \mathrm{e}^{-\lambda_r t} \phi_n^{(r)}, \tag{10.1.3}$$

where $\phi_n^{(r)}$ is the eigenvector corresponding to λ_r. It follows from the ordering of the eigenvalues that all eigenmodes $\phi_n^{(r)}$, $r > 0$, decay to zero much faster than the perturbed stationary density $\phi_n^{(0)}$. Thus at large times, we have the *quasistationary* approximation

$$p_n(t) \sim C_0 \mathrm{e}^{-\lambda_0 t} \phi_n^{(0)}. \tag{10.1.4}$$

Let T denote the (stochastic) FPT for which the system first reaches n_0, given that it started at $n = 0$, say. The distribution of FPTs is related to the survival probability $S(t)$ that the system hasn't yet reached n_0. That is, $\mathrm{Prob}\{t > T\} = S(t) = \sum_{n=0}^{n_0-1} p_n(t)$ and the FPT density is

$$f(t) = -\frac{dS}{dt} = -\sum_{n=0}^{n_0-1} \frac{dp_n(t)}{dt}. \tag{10.1.5}$$

Substituting for $p_n(t)$ using the quasistationary approximation (10.1.4) shows that the normalized FPT density can be approximated as $f(t) \sim \lambda_0 \mathrm{e}^{-\lambda_0 t}$ with λ_0^{-1} thus identified as the MFPT. In general, the eigenvalue equation for λ_0 and $\phi_n^{(0)}$ is difficult to analyze. On the other hand, one can use a WKB approximation to generate a quasistationary solution ϕ_n^{ε} for which $\mathbf{A}\boldsymbol{\phi}^{\varepsilon} = 0$ and $\phi_{n_0}^{\varepsilon} \sim O(\mathrm{e}^{-\eta N})$. Since the WKB solution does not satisfy the absorbing boundary condition, it is necessary to perform an asymptotic expansion in order to match the quasistationary solution

[1] The Perron–Frobenius theorem asserts that a real square matrix with positive entries has a unique largest real eigenvalue (the Perron eigenvalue) and that the corresponding eigenvector has strictly positive components. The theorem can also be extended to matrices with non-negative entries, provided that the matrix is irreducible. However, there can now be complex eigenvalues with the same absolute value as the Perron eigenvalue. In the case of a transition matrix, the Perron eigenvalue is zero. Strictly speaking, the Perron–Frobenius theorem applies to finite-dimensional matrices, so we will assume that it still holds in cases where the number of discrete states is infinite.

with the solution in a neighborhood of x_0. In the process this determines λ_0, whose inverse can be identified as the MFPT to escape from the basin of attraction of the metastable state x_- (see below). Dropping exponentially small terms, and writing $\phi_n^\varepsilon = \phi^\varepsilon(x)$ with x treated as a continuous variable, we have

$$\begin{aligned} 0 = {} & \Omega_+(x-1/N)\phi^\varepsilon(x-1/N) \\ & +\Omega_-(x+1/N)\phi^\varepsilon(x+1/N) - (\Omega_+(x)+\Omega_-(x))\phi^\varepsilon(x). \end{aligned} \tag{10.1.6}$$

We seek a WKB solution of the form

$$\phi^\varepsilon(x) \sim K(x;\varepsilon)\mathrm{e}^{-\Phi(x)/\varepsilon}, \tag{10.1.7}$$

with $K(x;\varepsilon) \sim \sum_{m=0}^{\infty} \varepsilon^m K_m(x)$. Substituting equation (10.1.7) into equation (10.1.6), Taylor expanding with respect to ε, and collecting the $O(1)$ terms give

$$\Omega_+(x)(\mathrm{e}^{\Phi'(x)}-1) + \Omega_-(x)(\mathrm{e}^{-\Phi'(x)}-1) = 0 \tag{10.1.8}$$

where $\Phi' = d\Phi/dx$. Solving this quadratic equation in $\mathrm{e}^{\Phi'}$ shows that

$$\Phi = \int^x \ln\frac{\Omega_-(y)}{\Omega_+(y)}dy \quad \text{or} \quad \Phi = \text{constant}. \tag{10.1.9}$$

Proceeding to the next level, equating terms at $\mathscr{O}(\varepsilon)$ gives

$$\Omega_+\mathrm{e}^{\Phi'}\left(-\frac{K_0'}{K_0}+\frac{\phi''}{2}\right) + \Omega_-\mathrm{e}^{-\Phi'}\left(\frac{K_0'}{K_0}+\frac{\phi''}{2}\right) - \Omega_+'\mathrm{e}^{\Phi'} + \Omega_-\mathrm{e}^{-\Phi'} = 0.$$

Substituting for Φ using (10.1.8) and solving for K_0 yields the following leading order forms for ϕ^ε:

$$\phi^\varepsilon(x) = \frac{A}{\sqrt{\Omega_+(x)\Omega_-(x)}}\mathrm{e}^{-N\Phi(x)}, \tag{10.1.10}$$

with Φ given by (10.1.9), which is sometimes called the activation solution, and

$$\phi^\varepsilon(x) = \frac{B}{\Omega_+(x)-\Omega_-(x)}, \tag{10.1.11}$$

which is sometimes called the relaxation solution. The constants A, B are determined by matching solutions around x_0. Clearly, (10.1.11) is singular at any fixed point x_j, where $\Omega_+(x_j) = \Omega_-(x_j)$, so is not a valid solution for the required quasistationary density. Note that the nontrivial WKB quasi-potential $\Phi(x)$ is identical to the one derived in equation (3.2.22).

Equation (10.1.8) has the form of a stationary *Hamilton–Jacobi equation* for Φ [306],

$$H(x, \Phi'(x)) = 0,$$

with *Hamiltonian*

$$H(x,p) = \sum_{r=\pm} \Omega_r(x)\,[\mathrm{e}^{rp} - 1]\,. \tag{10.1.12}$$

This suggests a corresponding classical mechanical interpretation, in which H determines the motion of a "particle" with position x and conjugate momentum p. A trajectory of the particle is given by the solution of Hamilton's equations

$$\dot{x} = \frac{\partial H}{\partial p} = \sum_{r=\pm 1} r\Omega_r(x)\mathrm{e}^{rp} \tag{10.1.13}$$

$$\dot{p} = -\frac{\partial H}{\partial x} = \sum_{r=\pm 1} \frac{\partial \Omega_r}{\partial x}(x)\,[1 - \mathrm{e}^{rp}]\,. \tag{10.1.14}$$

Here the time t should be viewed as a parameterization of paths rather than as a real time variable. In classical mechanics, Hamilton's equations can be derived from a variational principle, in which one looks for an extremum of the classical action

$$S[x,p] = \int_0^T [p\dot{x} - H(x,p)]dt$$

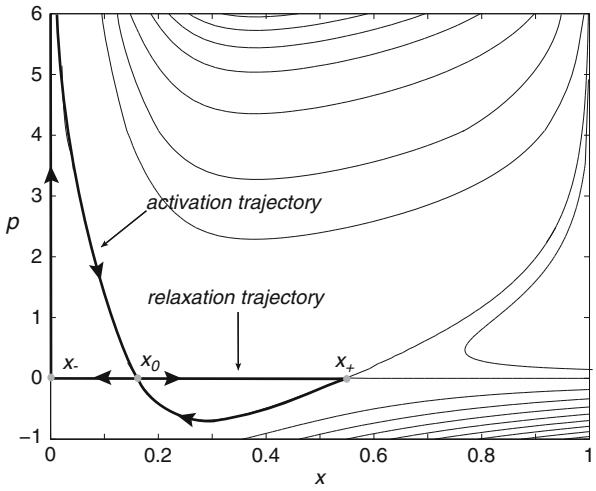


Fig. 10.2: Phase portrait of Hamiltonian equations of motion for transition rates $\Omega_\pm = \omega_\pm/N$ given by equation (3.4.14) of the Ca^{2+} sparks model, with $c_m = 0.01, c_e = 4, \alpha = 4$, and $k = 0.8$. The zero-energy solutions are shown as *thicker curves*

with respect to variations in x and p and assuming x is fixed at the endpoints, that is, $x(0) = \bar{x}$, $x(T) = x$. The corresponding quasi potential is given by

$$\Phi(x) = \inf_{x(0)=\bar{x},x(T)=x} S[x,p] \tag{10.1.15}$$

restricted to zero-energy solutions $H = 0$. Note that in classical mechanics one usually writes down a Lagrangian first, based on Newton's law of motion, and then constructs a Hamiltonian according to the Legendre transformation (see Box 10A and [306])

$$H(x,p) = p\cdot\dot{x} - L(x,\dot{x}), \quad p = \frac{\partial L}{\partial \dot{x}}. \tag{10.1.16}$$

Within the Lagrangian framework $\Phi(x)$ corresponds to the classical action evaluated along the least-action trajectory

$$\Phi(x) = \inf_{x(t_0)=\bar{x},x(T)=x} \int_0^T L(x,\dot{x})dt. \tag{10.1.17}$$

In terms of the underlying stochastic process $X(t)$, the least-action path is usually interpreted as the most probable fluctuational path from $\bar{x}$ to x (in the large N limit) [195, 409]. However, this cannot be justified within the WKB approximation, but must be established independently using a large deviation principle; we will show how to achieve this using path integrals. Assuming that such an interpretation is valid, the leading order term in the WKB approximation is determined by finding zero-energy solutions $p = p(x)$ such that $H(x,p(x)) = 0$. One solution is $p = 0$ or $\Phi =$ constant, which represents the classical action along a deterministic (or relaxation) trajectory. For example, once the system escapes from the metastable state x_-, it tends to rapidly converge to the other metastable state x_+ along such a deterministic path. (The contribution of relaxation trajectory to the mean escape time is usually neglected.) Another solution for Φ is (10.1.9), which can be interpreted as the action along a non-deterministic path that represents the most probable path of escape from x_- to x_0 [152, 163, 174]. In Fig. 10.2 we illustrate the Hamiltonian phase space for the Ca^{2+} sparks model of Sect. 3.4, showing the constant energy solutions of the Hamiltonian given by equation (10.1.12); the zero-energy activation and relaxation trajectories through the fixed points of the deterministic system are highlighted as thicker curves.

Box 10A. Least-action principle of classical mechanics.

One way to formulate the dynamics of a classical point particle with spatial coordinate $x(t)$ is in terms of the Lagrangian $L(x,\dot{x})$ and action [306]

$$S[x] = \int_{x(0)=x_0}^{x(T)=x_T} L(x,\dot{x})dt.$$

(For simplicity, we consider an autonomous system for which L is not explicitly dependent on time t.) If the particle has kinetic energy $m\dot{x}^2/2$ and potential energy $U(x)$, then

$$L(x,\dot{x}) = \frac{m}{2}\dot{x}^2 - U(x).$$

One can think of $S[x]$ as a *functional*, since it depends on the continuous set of values $x(t)$ over some time interval $t \in [0,T]$ with the endpoints fixed; the functional dependence is usually indicated by square brackets. The *least-action principle* states that the trajectories actually realized by the particle are those that extremize the action, $\delta S[x] = 0$. This means that, for any smooth function $y(t)$ with $y(0) = y(T) = 0$,

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}(S[x+\varepsilon y] - S[x]) = 0. \tag{10.1.18}$$

If we apply this definition to the action given by the time-integral of a Lagrangian, we have

$$\begin{aligned}\delta S[x] &= \int (L(x+\varepsilon y, \dot{x}+\varepsilon\dot{y}) - L(x,\dot{x}))dt \\ &= \varepsilon\int\left(\frac{\partial L(x,\dot{x})}{\partial x}y + \frac{\partial L(x,\dot{x})}{\partial \dot{x}}\dot{y}\right)dt + O(\varepsilon^2) \\ &= \varepsilon\int\left(\frac{\partial L(x,\dot{x})}{\partial x} - \frac{d}{dt}\frac{\partial L(x,\dot{x})}{\partial \dot{x}}\right)y dt + O(\varepsilon^2).\end{aligned}$$

The last step is obtained by performing an integration by parts and using the boundary conditions on y. Since $y(t)$ is arbitrary, the least action principle generates the *Euler–Lagrange equation*

$$\frac{\partial L(x,\dot{x})}{\partial x} = \frac{d}{dt}\frac{\partial L(x,\dot{x})}{\partial \dot{x}}.$$

Finally, substituting for L gives

$$m\ddot{x} = -\frac{dU(x)}{dx} \equiv F(x),$$

where $F(x)$ is the force generated by the potential $U(x)$. Thus the least-action principle is equivalent to Newton's law of motion.

Given the Lagrangian L, the corresponding classical Hamiltonian H is defined by the Legendre transformation

$$H(x,p) = p\dot{x} - L(x,\dot{x}).$$

with $p = \partial L/\partial\dot{x}$. Taking differential of $H = H(x,p)$ we have

$$dH = \frac{\partial H}{\partial x}dx + \frac{\partial H}{\partial p}dp.$$

Similarly, taking differentials of the Legendre transformation,

$$dH = \dot{x}dp + pd\dot{x} - \frac{\partial L}{\partial x}dx - \frac{\partial L}{\partial \dot{x}}d\dot{x}.$$

Using the definition of p to eliminate the $d\dot{x}$ terms and applying the Euler–Lagrange equation, we obtain Hamilton's equations

$$\dot{x} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial x}.$$

For the specific Lagrangian, $p = m\dot{x}$ (the momentum) and

$$H(x,p) = \frac{p^2}{2m} + U(x),$$

which is the total energy of the particle. Finally, if we allow the endpoint x of the action to vary, we obtain a function $\Phi(x)$ that satisfies a Hamilton–Jacobi equation. It is important to emphasize that the WKB method for a stochastic system works backward. The WKB solution of the quasistationary density yields a quasi-potential Φ that satisfies a Hamilton–Jacobi equation. From a mathematical perspective, there is an underlying Hamiltonian structure and an associated variational principle. However, the physical interpretation of the latter is not given a priori. Instead, one needs to apply some version of large deviation theory to identify solutions of the variational problem with most likely or optimal paths of the underlying stochastic process. This is established for the birth–death process using path integrals in Sect. 10.3.

Asymptotic Expansion Around x_0

Given the quasistationary approximation, the rate of escape from the metastable state centered about $x = x_-$ can be calculated by matching it with an appropriate inner solution in a neighborhood of the point $x = x_0$ [152, 163, 174, 252, 264]. This is necessary since the quasistationary solution (10.1.10) does not satisfy the absorbing boundary condition at the point x_0 separating the two metastable states. There are a number of different ways of carrying out the matched asymptotics; see

for example [264]. Here we will follow an approach based on fixing the probability flux J_0 through x_0 and then matching the activation solution for $x < x_0$ with the relaxation solution for $x > x_0$ using a diffusion approximation of the full master equation (10.3.1) in the vicinity of x_0 [163, 174, 252]. The latter yields the FPE (3.2.10), which can be rewritten in the form of a conservation equation

$$\frac{\partial}{\partial t}P(x,t) = -\frac{\partial}{\partial x}J(x,t) \tag{10.1.19}$$

with

$$J(x,t) = (\Omega_+(x) - \Omega_-(x))P(x,t) - \frac{1}{2N}\frac{\partial}{\partial x}\left[(\Omega_+(x) + \Omega_-(x))P(x,t)\right].$$

Substituting the quasistationary solution $p(x,t) = C_0 e^{-\lambda_0 t}\Pi(x)$ into equation (10.1.19) and using the fact that λ_0 is exponentially small gives

$$J_0 = (\Omega_+(x) - \Omega_-(x))\Pi(x) - \frac{1}{2N}\frac{\partial}{\partial x}\left[(\Omega_+(x) + \Omega_-(x))\Pi(x)\right],$$

where J_0 is the constant flux through x_0. In a neighborhood of x_0, this equation can be Taylor expanded to leading order in $x - x_0$ and integrated to obtain the solution

$$\Pi(x) = \frac{J_0 N}{\Omega_+(x_0)}\mathrm{e}^{(x-x_0)^2/\sigma^2}\int_x^\infty \mathrm{e}^{-(y-x_0)^2/\sigma^2}dy, \tag{10.1.20}$$

where

$$\sigma = \sqrt{\frac{2\Omega_+(x_0)}{N[\Omega_+'(x_0) - \Omega_-'(x_0)]}} \tag{10.1.21}$$

determines the size of the boundary layer around x_0.

In order to match the activation and relaxation solutions, the following asymptotic behavior of the inner solution (10.1.20) is used:

$$\Pi(x) = \begin{cases} \dfrac{NJ_0\sigma^2}{(x-x_0)\Omega_+(x_0)}, & x - x_0 \gg \sigma \\ \dfrac{NJ_0\sigma\sqrt{\pi}}{\Omega_+(x_0)}\mathrm{e}^{(x-x_0)^2/\sigma^2}, & x_0 - x \gg \sigma. \end{cases} \tag{10.1.22}$$

The solution to the right of the saddle matches the relaxation solution (10.1.11) since $\Omega_+(x) - \Omega_-(x) \approx (x-x_0)[\Omega_+'(x_0) - \Omega_-'(x_0)]$ for $x \approx x_0$ such that $B = J_0$. In order to match the solution on the left-hand side of x_0 with the activation solution (10.1.10), Taylor expand $\Phi(x)$ about x_0 using $\Phi'(x_0) = 0$ and $\Phi''(x_0) = 2/N\sigma^2$. It follows that

$$J_0 = \frac{A\Omega_+(x_0)}{\sqrt{\Omega_+(x_0)\Omega_-(x_0)}}\sqrt{\frac{|\Phi''(x_0)|}{2\pi N}}\mathrm{e}^{-N\Phi(x_0)}. \tag{10.1.23}$$

The final step in the analysis is to link the flux J_0 with the escape rate λ_0. This is achieved by substituting the quasistationary solution into the continuity equation (10.1.19) and integrating over the interval $x \in [0, x_0]$ with a reflecting boundary condition at $x = 0$:

$$\frac{1}{\lambda_0} = \frac{1}{J_0} \int_0^{x_0} \phi^{\varepsilon}(y) dy. \tag{10.1.24}$$

Since the activation solution is strongly peaked around the fixed point x_-, a Gaussian approximation of $\phi^{\varepsilon}(x)$ around x_- yields the final result (since $\Omega_+(x) = \Omega_-(x)$ at the fixed points x_0, x_-)

$$\lambda_0 = \frac{\Omega_+(x_-)}{2\pi} \sqrt{|\Phi''(x_0)|\Phi''(x_-)} \mathrm{e}^{-N[\Phi(x_0)-\Phi(x_-)]}. \tag{10.1.25}$$

Hence, we obtain equation (3.3.36) with $\tau_i = \lambda_0^{-1}$ and

$$\Phi''(x) = \frac{d}{dx} \ln \left(\frac{\Omega_-(x)}{\Omega_+(x)} \right) = \frac{\Omega_-'(x)}{\Omega_-(x)} - \frac{\Omega_+'(x)}{\Omega_+(x)} = \gamma(x).$$

10.1.2 WKB Method for a Stochastic Hybrid System

In Sect. 3.5 we considered a model of membrane voltage fluctuations that involved the coupling between a piecewise deterministic dynamical system and a jump Markov process, which is known as a stochastic hybrid system or piecewise deterministic Markov process [135]. Other examples of stochastic hybrid systems considered in this book include bacterial chemotaxis (Sect. 5.3), stochastic gene networks (Chap. 6), and motor-driven intracellular transport (Sect. 7.4). All of these systems can be modeled in terms of an equation having the form

$$\frac{dx}{dt} = \frac{1}{\tau_x} F_n(x) \tag{10.1.26}$$

where $x \in \mathbb{R}^d$ represents d continuous variables, $n \in I \subseteq \mathbb{Z}$ is a discrete internal state variable, and the latter evolves according to a jump Markov process $n' \to n$ with transition rates $W_{nn'}(x)/\tau_n$. For simplicity, we restrict ourselves to a single continuous variable ($d = 1$). (It is also possible to have a set of discrete variables, but one can always relabel the internal states so that they are effectively indexed by a single integer.) The jump propagator $W_{nn'}(x)dt/\tau_n$ is the probability that the system at x switches from the discrete internal state n' at time t to the state n at time $t + dt$. The transition rates generally depend on x, with the latter coupled to the associated jump Markov process according to equation (10.1.26), which is only defined between jumps, during which $x(t)$ evolves deterministically. Denote the random state of the full model (10.1.26) at time t by $(X(t), N(t))$, and introduce the corresponding probability density

$$\mathrm{Prob}\{X(t) \in (x, x+dx), N(t) = n\} = p_n(x,t)dx, \tag{10.1.27}$$

given initial conditions $X(0) = x_0, N(0) = n_0$ (which are suppressed). The probability density evolves according to the differential Chapman–Kolmogorov (CK) equation (dropping the explicit dependence on initial conditions)

$$\frac{\partial p}{\partial t} + \frac{1}{\tau_x}\frac{\partial [F_n(x)p_n(x,t)]}{\partial x} = \frac{1}{\tau_n}\sum_{n'\in I}[W_{nn'}(x)p_{n'}(x,t) - W_{n'n}(x)p_n(x,t)]\,. \tag{10.1.28}$$

We have introduced two time scales in the system, a relaxation time scale τ_x for the x-dynamics and a transition time scale τ_n for the jump process.

In many of the listed biophysical applications, the kinetics associated with the jump process are much faster than the relaxation dynamics of x, that is, $\tau_n \ll \tau_x$. Let us fix the time units by setting $\tau_x = 1$ and introduce the small parameter $\varepsilon = \tau_n/\tau_x$. We can then rewrite (10.1.28) in the more compact form

$$\frac{\partial p_n}{\partial t} = -\frac{\partial [F_n(x)p_n(x,t)]}{\partial x} + \frac{1}{\varepsilon}\sum_{n'\in I} A_{nn'}(x)p(x), \tag{10.1.29}$$

with

$$A_{nn'}(x) = W_{nn'}(x) - \sum_{m\in I} W_{mn}(x)\delta_{n',n}.$$

In the limit $\varepsilon \to 0$, equation (10.1.26) reduces to the deterministic or mean-field equation

$$\frac{dx}{dt} = \mathscr{F}(x) \equiv \sum_{n\in I} F_n(x)\rho_n(x), \tag{10.1.30}$$

where $\rho_n(x)$ is the unique steady-state density satisfying $\sum_{m\in I} A_{nm}(x)\rho_m(x) = 0$. This follows from the law of large numbers [177]. As in the case of the birth–death master equation, we are assuming that for fixed x, the matrix $A_{nm}(x)$ is irreducible and has a simple zero eigenvalue with corresponding left eigenvector $\mathbf{1}$ whose components are all unity, that is, $\sum_{n\in I} A_{nm}(x) = 0$ for all m. The Perron–Frobenius theorem then ensures that all other eigenvalues are negative and the continuous-time Markov process for fixed x,

$$\frac{dp_n(x,t)}{dt} = \frac{1}{\varepsilon}\sum_{m\in I} A_{nm}(x)p_m(x,t),$$

has a globally attracting steady state $\rho_n(x)$ such that $p_n(x,t) \to \rho_n(n)$ as $t \to \infty$.

Following along similar lines to the analysis of the birth–death master equation, suppose that the mean-field equation (10.1.30) is bistable with a pair of stable fixed points $x_\pm$ separated by an unstable fixed point x_0. Assume that the stochastic system is initially at x_-. On short time scales ($t \ll 1/\varepsilon$) the system rapidly converges to a quasistationary solution within the basin of attraction of x_-, which can be approximated by a Gaussian solution of the reduced FP equation obtained using a QSS diffusion or adiabatic approximation of the CK equation (10.1.29) (see Sect. 7.4).

However, on longer time scales, the survival probability slowly decreases due to rare transitions across x_0 at exponentially small rates, which cannot be calculated accurately using the QSS diffusion approximation. One thus has to work with the full CK equation (10.1.29) supplemented by an absorbing boundary condition at x_0:

$$p_n(x_0,t) = 0, \text{ for all } n \in \Sigma, \tag{10.1.31}$$

where $\Sigma \subset I$ is the set of internal states n for which $F_n(x_0) < 0$. The initial condition is taken to be

$$p_n(x,0) = \delta(x - x_-)\delta_{n,\bar{n}}. \tag{10.1.32}$$

Let T denote the (stochastic) FPT for which the system first reaches x_0, given that it started at x_-. The distribution of FPTs is related to the survival probability that the system hasn't yet reached x_0, that is,

$$\text{Prob}\{t > T\} = S(t) \equiv \int_{-\infty}^{x_0} \sum_{n\in I} p_n(x,t)dx.$$

The FPT density is then

$$f(t) = -\frac{dS}{dt} = -\int_{-\infty}^{x_0} \sum_{n\in I} \frac{\partial p_n}{\partial t}(x,t)dx. \tag{10.1.33}$$

Substituting for $\partial p_n/\partial t$ using the CK equation (10.1.29) shows that

$$f(t) = \int_{-\infty}^{x_0} \left[\sum_{n\in I} \frac{\partial [F_n(x)p_n(x,t)]}{\partial x}\right] dx = \sum_{n\in I} p_n(x_0,t)F_n(x_0). \tag{10.1.34}$$

We have used $\sum_n A_{nm}(x) = 0$ and $\lim_{x\to-\infty} F_n(x)p_n(x,t) = 0$. The FPT density can thus be interpreted as the probability flux $J(x,t)$ at the absorbing boundary, since we have the conservation law

$$\sum_{n\in I} \frac{\partial p_n(x,t)}{\partial t} = -\frac{\partial J(x,t)}{\partial x}, \quad J(x,t) = \sum_{n\in I} F_n(x)p_n(x,t). \tag{10.1.35}$$

Again proceeding as in the analysis of the birth–death master equation, suppose that the solution to the CK equation (10.1.29) with absorbing boundary at $x = x_0$ has the eigenfunction expansion (see Sect. 2.5)

$$p_n(x,t) = \sum_{r\geq 0} C_r e^{-\lambda_r t}\phi_n^{(r)}(x), \tag{10.1.36}$$

where $\text{Re}[\lambda_r] > 0$ for all $r \geq 0$ and $(\lambda_r, \phi_n^{(r)})$ are an eigenpair of the linear operator $\widehat{L}$ on the right-hand side of the CK equation (10.1.29):

$$\widehat{L}\phi_n^{(r)}(x) \equiv \frac{d}{dx}(F_n(x)\phi_n^{(r)}(x)) - \frac{1}{\varepsilon}\sum_{m\in I} A_{nm}\phi_m^{(r)}(x) = \lambda_r\phi_n^{(r)}(x), \tag{10.1.37}$$

together with the boundary conditions

$$\phi_n^{(r)}(x_0) = 0, \text{for} n \in \Sigma. \tag{10.1.38}$$

Furthermore, assume that there exists a simple, real eigenvalue λ_0 such that $0 < \lambda_0 < \text{Re}[\lambda_1] \leq \text{Re}[\lambda_2] \leq \ldots$ with $\lambda_0 \sim \text{e}^{-C/\varepsilon}$, whereas $\text{Re}[\lambda_r] = O(1)$ for $r > 0$. It follows that at large times we have the quasistationary approximation

$$p_n(x,t) \sim C_0 \text{e}^{-\lambda_0 t} \phi_n^{(0)}(x). \tag{10.1.39}$$

Substituting such an approximation into equation (10.1.34) implies that $f(t) \sim \lambda_0 \text{e}^{-\lambda_0 t}$ with λ_0^{-1} the MFPT. As shown elsewhere [321, 475], one can determine λ_0 by first using a WKB approximation to construct a quasistationary solution ϕ_n^ε for which $\widehat{L}\phi^\varepsilon = 0$ and $\phi_n^\varepsilon(x_0) \sim O(\text{e}^{-\eta N})$ and then performing an asymptotic expansion in order to match the quasistationary solution with the solution in a neighborhood of x_0. Here we will simply calculate the quasi-potential arising from the leading order terms in the WKB approximation $\phi_n^\varepsilon(x)$. The latter takes the form

$$\phi_n^\varepsilon(x) \sim R_n(x) \exp\left(-\frac{\Phi(x)}{\varepsilon}\right), \tag{10.1.40}$$

where $\Phi(x)$ is the quasi-potential. Substituting into the time-independent version of equation (10.1.29) yields

$$\sum_{\mathbf{m}\in I} \left(A_{nm}(x) + \Phi'(x)\delta_{\mathbf{n},\mathbf{m}} F_m(x)\right) R_m(x) = \varepsilon \frac{dF_n(x)R_n(x)}{dx}, \tag{10.1.41}$$

where $\Phi' = d\Phi/dx$. Introducing the asymptotic expansions $R \sim R^{(0)} + \varepsilon R^{(1)}$ and $\Phi \sim \Phi_0 + \varepsilon\Phi_1$, the leading order equation is

$$\sum_{m\in I} A_{nm}(x) R_m^{(0)}(x) = -\Phi_0'(x) F_n(x) R_n^{(0)}(x). \tag{10.1.42}$$

Positivity of the probability density ϕ_ε requires positivity of the corresponding solution $R^{(0)}$. One positive solution is the trivial solution $R^{(0)} = \rho$, for which $\Phi_0' = 0$. It can be proven that if $F_n(x)$ for fixed $x \in \Omega$ changes sign as n increases from zero, then there exists one other positive eigenfunction $R^{(0)}$, which can be identified as the appropriate WKB solution [475].

There are two major differences between the WKB analysis of the stochastic hybrid system and the birth–death master equation

(i) In the former case the small parameter ε arises from the separation of timescales, whereas in the latter case $\varepsilon = N^{-1}$ where N is the system size. Thus equation (10.1.40) holds irrespective of the number of molecules N.

(ii) The quasi-potential of the birth–death master equation naturally satisfies a Hamilton–Jacobi equation (see (10.1.8)), whereas the quasi-potential of the stochastic hybrid system is related to the eigenvalue of a matrix operator (see (10.1.42)).

It turns out that there is a Hamiltonian structure underlying the construction of the quasi-potential for stochastic hybrid systems, which can be shown rigorously using a large deviation principle [332]. In Sect. 10.4 we derive such an action principle using path-integrals along the lines of [66]. Here we simply summarize the main result, namely that one can identify the Hamiltonian $H(x,p)$ as the unique Perron (principle) eigenvalue $\Lambda_0(x,p)$ of the matrix equation

$$\sum_m [A_{nm}(x) + p\delta_{n,m}F_m(x)]\, R_m^{(0)}(x,p) = \Lambda_0(x,p) R_n^{(0)}(x,p). \tag{10.1.43}$$

Comparison with equation (10.1.42) show, that the quasi-potential is obtained from the Hamilton–Jacobi equation $\Lambda_0(x,\Phi'(x)) = 0$. This is equivalent to finding zero-energy solutions of Hamilton's equations

$$\dot{x} = \frac{\partial \Lambda_0(x,p)}{\partial p}, \quad \dot{p} = -\frac{\partial \Lambda_0(x,p)}{\partial x}, \tag{10.1.44}$$

and identifying Φ_0 as the action along the resulting solution curve $(x(t),p(t))$:

$$\Phi_0(x) = \int_{\bar{x}}^{x} p(x')dx'. \tag{10.1.45}$$

10.1.3 Stochastic Ion Channels Revisited

We will illustrate the above analysis by returning to the stochastic conductance-based model given by equations (3.5.13) and (3.5.14), which was analyzed in Sect. 3.5 using a diffusion approximation of the underlying CK equation (3.5.17) for large N and small ε (fast Na channels). It is convenient to rewrite the CK equation (3.5.17) in the form (10.1.29) with

$$F_n(x) \equiv \frac{1}{N} f(x) n - g(x)$$

and A the tridiagonal matrix (for fixed x):

$$A_{n,n-1;x} = \omega_+(x,n-1), A_{n,n;x} = -\omega_+(x,n) - \omega_-(n), A_{n,n+1;x} = \omega_-(n+1) \tag{10.1.46}$$

for $n = 0,1,\ldots,N$. In the mean-field limit $\varepsilon \to 0$, we recover the deterministic kinetic equation (3.3.24),

$$\frac{dx}{dt} = a(x)f(x) - g(x) \tag{10.1.47}$$

where

$$a(x) = \langle n \rangle / N, \quad \langle n \rangle = \sum_{n=1}^{N} n\rho_n(x),$$

and ρ is the quasistationary density

$$\rho_n(x) = \frac{N!}{(N-n)!n!} a(x)^n b(x)^{N-n}, \quad a(x) = \frac{\alpha(x)}{\alpha(x)+\beta(x)}, b(x) = \frac{\beta(x)}{\alpha(x)+\beta(x)}. \tag{10.1.48}$$

In the case of the stochastic ion channel model, equation (10.1.43) with $R_n^{(0)}(x,p) = \psi_n(x,p)$ takes the explicit form

$$\begin{aligned}(N-n+1)\alpha\psi_{n-1} - [\lambda_0 + n\beta + (N-n)\alpha]\psi_n \\ +(n+1)\beta\psi_{n+1} = -p\left(\frac{n}{N}f - g\right)\psi_n.\end{aligned} \tag{10.1.49}$$

Consider the trial solution

$$\psi_n(x,p) = \frac{\Gamma(x,p)^n}{(N-n)!n!}, \tag{10.1.50}$$

which yields the following equation relating Γ and Λ_0:

$$\frac{n\alpha}{\Gamma} + \Gamma\beta(N-n) - \Lambda_0 - n\beta - (N-n)\alpha = -p\left(\frac{n}{N}f - g\right).$$

Collecting terms independent of n and terms linear in n yields the pair of equations

$$p = -\frac{N}{f(x)}\left(\frac{1}{\Gamma(x,p)} + 1\right)(\alpha(x) - \beta(x)\Gamma(x,p)) \tag{10.1.51}$$

and

$$\Lambda_0(x,p) = -N(\alpha(x) - \Gamma(x,p)\beta(x)) - pg(x). \tag{10.1.52}$$

Eliminating Γ from these equation gives

$$p = \frac{1}{f(x)}\left(\frac{N\beta(x)}{\Lambda_0(x,p) + N\alpha(x) + pg(x)} + 1\right)(\Lambda_0(x,p) + pg(x)).$$

This yields a quadratic equation for Λ_0 of the form

$$\Lambda_0^2 + \sigma(x)\Lambda_0 - h(x,p) = 0 \tag{10.1.53}$$

with

$$\begin{aligned}\sigma(x) &= (2g(x) - f(x)) + N(\alpha(x) + \beta(x)), \\ h(x,p) &= p[-N\beta(x)g(x) + (N\alpha(x) + pg(x))(f(x) - g(x))].\end{aligned}$$

Along the zero-energy surface $\Lambda_0(x,p) = 0$, we have $h(x,p) = 0$ which yields the pair of solutions

$$p = 0 \text{ and } p = -\mu(x) \equiv N\frac{\alpha(x)f(x) - (\alpha(x)+\beta)g(x)}{g(x)(f(x)-g(x))}. \tag{10.1.54}$$

It follows that the nontrivial WKB quasi-potential is given by

$$\Phi_0(x) = -\int^x \mu(y)dy, \tag{10.1.55}$$

which differs significantly from the FP quasi-potential $\Phi_{\rm FP}$ of equation (3.5.26). This is a major source of the error in the diffusion approximation illustrated in Fig. 3.18. Further examples of WKB analysis are considered in Exs. 10.1 and 10.2.

10.2 Path Integral Representation of an SDE

Consider the scalar SDE

$$dX(t) = A(X)dt + \sqrt{\varepsilon}dW(t), \tag{10.2.1}$$

for $0 \le t \le T$ and initial condition $X(0) = x_0$. Here $W(t)$ is a Wiener process and the noise is taken to be weak ($\varepsilon \ll 1$). Discretizing time by dividing the interval $[0,T]$ into N equal subintervals of size Δt such that $T = N\Delta t$ and setting $X_n = X(n\Delta t)$, we have

$$X_{n+1} - X_n = A(X_n)\Delta t + \sqrt{\varepsilon}\Delta W_n,$$

with $n = 0, 1, \ldots, N-1$, $\Delta W_n = W((n+1)\Delta t) - W(n\Delta t)$

$$\langle \Delta W_n \rangle = 0, \quad \langle \Delta W_m \Delta W_n \rangle = \Delta t \delta_{m,n}.$$

Let $\mathbf{X}$ and $\mathbf{W}$ denote the vectors with components X_n and W_n, respectively. Formally, the conditional probability density function for $\mathbf{X} = \mathbf{x}$ given a particular realization $\mathbf{w}$ of the stochastic process $\mathbf{W}$ (and initial condition x_0) is

$$P(\mathbf{x}|\mathbf{w}) = \prod_{n=0}^{N-1} \delta\left(x_{n+1} - x_n - A(x_n)\Delta t - \sqrt{\varepsilon}\Delta w_n\right).$$

Inserting the Fourier representation of the Dirac delta function,

$$\delta(x_{m+1} - z_m) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{-i\tilde{x}_m(x_{m+1}-z_m)} d\tilde{x}_m, \tag{10.2.2}$$

gives

$$P(\mathbf{x}|\mathbf{w}) = \prod_{m=0}^{N-1}\left[\int_{-\infty}^{\infty} \mathrm{e}^{-i\tilde{x}_m\left(x_{m+1} - x_m - A(x_m)\Delta t - \sqrt{\varepsilon}\Delta w_m\right)} \frac{d\tilde{x}_m}{2\pi}\right].$$

The Gaussian random variable ΔW_n has the probability density function

$$P(\Delta w_n) = \frac{1}{\sqrt{2\pi\Delta t}} \mathrm{e}^{-\Delta w_n^2/2\Delta t}.$$

Hence, setting

$$P(\mathbf{x}) = \int P[\mathbf{x}|\mathbf{w}] \prod_{n=0}^{N-1} P(\Delta w_n) d\Delta w_n$$

and performing the integration with respect to Δw_n by completing the square, we obtain the result

$$P(\mathbf{x}) = \prod_{m=0}^{N-1} \left[\int_{-\infty}^{\infty} \mathrm{e}^{-i\tilde{x}_m (x_{m+1} - x_m - A(x_m)\Delta t)} \mathrm{e}^{-\varepsilon \tilde{x}_m^2 \Delta t/2} \frac{d\tilde{x}_m}{2\pi} \right].$$

Finally, performing the Gaussian integration with respect to $\tilde{x}_m$, we have

$$\begin{aligned} P(\mathbf{x}) &= \prod_{m=0}^{N-1} \frac{1}{\sqrt{2\pi\varepsilon\Delta t}} \mathrm{e}^{-(x_{m+1} - x_m - A(x_m)\Delta t)^2/(2\varepsilon\Delta t)} \\ &= \mathscr{N} \exp\left[-\sum_{m=0}^{N-1} (x_{m+1} - x_m - A(x_m)\Delta t)^2/(2\varepsilon\Delta t) \right] \\ &= \mathscr{N} \exp\left[-\frac{1}{2\varepsilon} \sum_{m=0}^{N-1} \left(\frac{x_{m+1} - x_m}{\Delta t} - A(x_m) \right)^2 \Delta t \right], \end{aligned} \tag{10.2.3}$$

with

$$\mathscr{N} = \frac{1}{(2\pi\varepsilon\Delta t)^{N/2}}.$$

Note that $P(\mathbf{x}) dx_1 \ldots dx_N$ is the probability that a given realization $X(t)$ of the discretized stochastic process lies within an infinitesimal domain given by $x_n < X(n\Delta t) < x_n + dx_n$ for $n = 1, \ldots N$ and with initial condition $X(0) = x_0$. One can define expectations in the usual way with

$$\mathbb{E}[F(\mathbf{X})] = \int F(\mathbf{x}) P(\mathbf{x}) dx_1 \ldots x_N$$

for any integrable function F. The main value of this particular integral representation is that it has a well-defined continuum limit obtained by taking $\Delta t \to 0, N \to \infty$ with $N\Delta t = T$ fixed. Now $P[x]$ is a probability density *functional* over the different paths $\{x(t)\}_0^T$ realized by the original SDE (10.2.1) with $X(0) = x_0$. Taking the continuum limit of the exponential in equation (10.2.3) shows that (up to an appropriate normalization)

$$P[x] \sim \exp\left[-\frac{1}{2\varepsilon} \int_0^T (\dot{x} - A(x))^2 dt \right], \tag{10.2.4}$$

and the expectation of a functional $F[x]$ is given by the Onsager–Machlup path integral [235, 236]

$$\mathbb{E}[F[x]] = \int F[x] P[x] \mathscr{D}(x), \tag{10.2.5}$$

where the probability measure $\mathscr{D}[x]$ on the space of trajectories is known as the Wiener measure. (It turns out that the subtleties of measure theory are not needed in

order to carry out useful calculations using path-integrals.) Suppose, for example, one wants to determine the probability density

$$P(x,\tau|x_0) = \int_{x(0)=x_0}^{x(\tau)=x} P[x]\mathscr{D}[x]$$

that the stochastic process $X(t)$ reaches a point x at time $t=\tau$ given that it started at x_0 at time $t=0$. Substituting the exponential form for $P[x]$, we see that

$$P(x,\tau|x_0) = \int_{x(0)=x_0}^{x(\tau)=x} \exp\left[-\frac{1}{2\varepsilon}\int_0^\tau (\dot{x}-A(x))^2 dt\right]\mathscr{D}[x]. \tag{10.2.6}$$

In the limit $\varepsilon \to 0$, we can use the method of steepest descents to obtain the approximation

$$P(x,\tau|x_0) \sim \exp\left[-\frac{\Phi(x,\tau|x_0)}{\varepsilon}\right], \tag{10.2.7}$$

where Φ is known as the stochastic or quasi-potential

$$\Phi(x,\tau|x_0) = \inf_{x(0)=x_0, x(\tau)=x} S[x], \tag{10.2.8}$$

with

$$S[x] = \int_0^\tau L(x,\dot{x})dt \tag{10.2.9}$$

and

$$L(x,\dot{x}) = \frac{1}{2}(\dot{x}-A(x))^2. \tag{10.2.10}$$

An interesting feature of the above analysis is that the quasi-potential is obtained by solving a variational problem that minimizes the functional $S[x]$ over trajectories from $\{x(t)\}_0^\tau$ with $x(0)=x_0$ and $x(\tau)=x$. In other words, the variational problem determines the most probable path. If we now make the connection with classical mechanics, we can identify $S[x]$ as an action with corresponding Lagrangian $L(x,\dot{x})$, and the most probable path is given by the solution to the Euler–Lagrange equation (see Box 10A)

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{x}} = \frac{\partial L}{\partial x}. \tag{10.2.11}$$

Substituting for L, we see that the most probable path satisfies

$$\ddot{x} - A'(x)\dot{x} = -(\dot{x}-A(x))A'(x),$$

that is

$$\ddot{x} = A(x)A'(x). \tag{10.2.12}$$

Suppose that in the zero-noise limit there is a globally attracting fixed point x_s such that $A(x_s)=0$. The steady-state solution of the corresponding FP equation can be obtained by solving the Euler–Lagrange equation with the conditions $x(-\infty)=x_s$

and $x(\tau) = x$. Multiplying both sides of equation (10.2.12) by $\dot{x}$ and integrating with respect to t shows that $\dot{x}(t)^2 = A(x(t))^2 + \text{constant}$. The initial condition implies that the constant is zero and the end condition implies that the most probable path satisfies $\dot{x} = -A(x)$. It follows that the quasi-potential is

$$\Phi(x,\tau|x^*) = -2\int_{-\infty}^{\tau} A(x)\dot{x}dt = 2\int_{-\infty}^{\tau} U'(x)\dot{x}dt = 2\int_{x_s}^{x} U'(x)dx = 2U(x),$$

where we have set $A(x) = -U'(x)$ with $U(x)$ the potential of the deterministic system. Hence, we obtain the expected result that the stationary density is

$$P(x) \sim \mathrm{e}^{-2U(x)/\varepsilon}.$$

In the case of a scalar SDE one can of course solve the steady-state FP equation directly. The power of the path-integral formulation is that one can extend the definition of the quasi-potential to multivariate SDEs and to nonlinear systems having multiple attractors. For example, consider the multivariate SDE

$$dX_i(t) = A_i(\mathbf{X})dt + \sqrt{\varepsilon}\sum_j b_{ij}(\mathbf{X})dW_i(t), \tag{10.2.13}$$

for $i = 1,\ldots,d$ with $W_i(t)$ a set of independent Wiener processes. Generalizing the path-integral method to higher dimensions or using the rigorous approach of Freidlin and Wentzel [195], one obtains a large deviation principle with action functional

$$S[\mathbf{x}] = \frac{1}{2}\int_0^T \sum_{i,j=1}^{d} (\dot{x}_i(t) - A_i(\mathbf{x}(t)))D_{ij}^{-1}(\dot{x}_j(t) - A_j(\mathbf{x}(t)))dt, \tag{10.2.14}$$

where $\mathbf{D} = \mathbf{b}\mathbf{b}^{\mathrm{tr}}$ is the diffusion matrix. Suppose that the underlying deterministic system has multiple attracting fixed points $\mathbf{x}_r, r = 1,2\ldots$. The quasi-potential $\Phi(\mathbf{x})$ characterizing the stationary distribution (assuming it exists) is estimated as $\Phi(\mathbf{x}) = \min_r \Phi_r(\mathbf{x})$, where $\Phi_r(\mathbf{x})$ is the quasi-potential obtained by initiating paths at $\mathbf{x}_r$:

$$\Phi_r(\mathbf{x}) = \inf_{\mathbf{x}(-\infty)=\mathbf{x}_r, \mathbf{x}(\tau)=\mathbf{x}} S[\mathbf{x}].$$

Note that the minimum over r switches abruptly on a separatrix separating the basins of attraction of the fixed points.

10.2.1 The WKB Method and First Passage Time Problems in the Weak Noise Limit

We now establish a connection to the variational principle obtained from path integrals with a Hamiltonian formalism derived using the WKB method. The latter was applied to a jump Markov process in Sect. 10.1.1 and a stochastic hybrid system in

Sect. 10.1.2. Consider the FP equation corresponding to the scalar SDE (10.2.1):

$$\frac{\partial p}{\partial t} = -\frac{\partial [A(x)p(x,t)]}{\partial x} + \frac{\varepsilon}{2}\frac{\partial^2 p(x,t)}{\partial x^2} \equiv -\frac{\partial J(x,t)}{\partial x}, \tag{10.2.15}$$

where

$$J(x,t) = -\frac{\varepsilon}{2}\frac{\partial p(x,t)}{\partial x} + A(x)p(x,t).$$

Suppose that the deterministic equation $\dot{x} = A(x)$ has a stable fixed point x_-, $A(x_-) = 0$, and a basin of attraction given by the interval $\Omega = (0, x_0)$; the point x_0 corresponds to an unstable fixed point. For small but finite ε the fluctuations about the steady state $p(x)$ can induce rare transitions out of the basin of attraction due to a metastable trajectory crossing the point x_0. Assume that the stochastic system is initially at x_- so that $p(x,0) = \delta(x - x_-)$. In order to solve the FPT problem for escape from the basin of attraction of x_-, we impose an absorbing boundary condition at x_0, $p(x_0,t) = 0$ and a reflecting boundary condition at $x = 0$. Let T denote the (stochastic) FPT for which the system first reaches x_0, given that it started at x_-. The distribution of FPTs is related to the survival probability that the system hasn't yet reached x_0:

$$S(t) \equiv \int_\Omega p(x,t)dx. \tag{10.2.16}$$

That is, $\text{Prob}\{t > T\} = S(t)$ and the FPT density is

$$f(t) = -\frac{dS}{dt} = -\int_\Omega \frac{\partial p}{\partial t}(x,t)dx. \tag{10.2.17}$$

Substituting for $\partial p/\partial t$ using the FP equation (10.2.15) shows that

$$f(t) = \int_\Omega \frac{\partial J(x,t)}{\partial x}dx = J(x_0,t) = -\frac{\varepsilon}{2}\frac{\partial p(x_0,t)}{\partial x}. \tag{10.2.18}$$

We have used $J(0,t) = 0$ and $p(x_0,t) = 0$. The FPT density can thus be interpreted as the probability flux $J(x,t)$ at the absorbing boundary.

The FPT problem in the weak noise limit ($\varepsilon \ll 1$) has been well studied in the case of FP equations; see for example [409, 426, 466, 574]. One of the characteristic features of the weak noise limit is that the flux through the absorbing boundary and the inverse of the MFPT $\langle T \rangle$ are exponentially small, that is, $\langle T \rangle \sim \mathrm{e}^{C/\varepsilon}$ for some constant C (reflecting an underlying large deviation principle). In order to make this connection more explicit, we will proceed along analogous lines to Sect. 10.1 and consider the eigenfunction expansion

$$p(x,t) = \sum_r C_r \mathrm{e}^{-\lambda_r t}\phi_r(x), \tag{10.2.19}$$

where $(\lambda_r, \phi_r(x))$ are an eigenpair of the linear operator

$$\widehat{L} = \frac{\partial}{\partial x} A(x) - \frac{\varepsilon}{2} \frac{\partial^2}{\partial x^2}$$

appearing on the right-hand side of (10.2.15). That is,

$$\widehat{L}\phi_r(x) = \lambda_r \phi_r(x), \tag{10.2.20}$$

together with the absorbing boundary conditions $\phi_r(x_0) = 0$. We also assume that the eigenvalues λ_r all have positive definite real parts and the smallest eigenvalue λ_0 is real and simple, so that we can introduce the ordering $0 < \lambda_0 < \mathrm{Re}[\lambda_1] \leq \mathrm{Re}[\lambda_2] \leq \ldots$. The exponentially slow rate of escape through x_0 in the weak noise limit means that λ_0 is exponentially small, $\lambda_0 \sim \mathrm{e}^{-C/\varepsilon}$, whereas $\mathrm{Re}[\lambda_r] = O(1)$ for $r \geq 1$. Under the above assumptions, we have the *quasistationary* approximation for large t

$$p(x,t) \sim C_0 \mathrm{e}^{-\lambda_0 t} \phi_0(x), \tag{10.2.21}$$

and the FPT density takes the form $f(t) \sim \lambda_0 \mathrm{e}^{-\lambda_0 t}$ with λ_0^{-1} identified as the MFPT.

As in the examples considered in Sect. 10.1, the calculation of the principle eigenvalue λ_0 consists of two major components [409, 426, 466, 574]: (i) a WKB approximation of the quasistationary state, which also provides an alternative method for deriving the quasi-potential of large deviation theory, and (ii) the use of matched asymptotics in order to match the outer quasistationary solution with an inner solution within a boundary layer around x_0 so that the absorbing boundary condition is satisfied. Here we will focus on the WKB method for determining the quasi-potential. We seek a quasistationary solution of the WKB form

$$\phi^{\varepsilon}(x) \sim K(x;\varepsilon) \mathrm{e}^{-\Phi(x)/\varepsilon}, \tag{10.2.22}$$

with $K(x;\varepsilon) \sim \sum_{m=0}^{\infty} \varepsilon^m K_m(x)$. Substitute equation (10.2.22) into the eigenvalue equation $\widehat{L}\phi_0(x) = \lambda_0 \phi_0(x)$ and Taylor expand with respect to ε using the fact that λ_0 is exponentially small. Collecting the $O(1)$ terms gives

$$\frac{1}{2}\left(\frac{\partial \Phi(x)}{\partial x}\right)^2 + A(x)\frac{\partial \Phi(x)}{\partial x} = 0. \tag{10.2.23}$$

Similarly, collecting $O(\varepsilon)$ terms yields the following equation for the leading contribution K_0 to the pre factor:

$$\left[\frac{\partial \Phi}{\partial x} + A(x)\right]\frac{\partial K_0}{\partial x} = -\left[A'(x) + \frac{1}{2}\frac{\partial^2 \Phi(x)}{\partial x^2}\right] K_0(x). \tag{10.2.24}$$

Equation (10.2.23) has the form of a Hamilton–Jacobi (HJ) equation for a classical Newtonian particle. That is, introducing the time-independent Hamiltonian

$$H(x,p) = \frac{p^2}{2} + A(x)p, \tag{10.2.25}$$

we see that equation (10.2.23) can be rewritten as the "zero-energy" HJ equation

$$H(x, \Phi'(x)) = 0. \tag{10.2.26}$$

The HJ structure suggests a classical mechanical interpretation, in which the Hamiltonian H describes the motion of a "fictitious" particle with position x and conjugate momentum p evolving according to Hamilton's equations

$$\dot{x} = \frac{\partial H}{\partial p} = p + A(x) \tag{10.2.27a}$$

$$\dot{p} = -\frac{\partial H}{\partial x} = -pA'(x). \tag{10.2.27b}$$

The Hamiltonian is related to a classical Lagrangian $L(x,\dot{x})$ according to the Legendre transformation

$$H(x,p) = p\dot{x} - L(x,\dot{x}), \quad p = \frac{\partial L}{\partial \dot{x}}.$$

It follows that

$$L(x,\dot{x}) = \frac{1}{2}(\dot{x} - A(x))^2, \tag{10.2.28}$$

which we immediately recognize as the Lagrangian associated with the action functional of the large deviation principle (10.2.9). From the least-action principle of classical mechanics, one knows that $\Phi(x)$ is given by

$$\Phi(x) = \inf_{x(t_0)=x^*, x(\tau)=x} \int_0^\tau L(x,\dot{x})dt, \tag{10.2.29}$$

and can thus be identified as the quasi-potential of large deviation theory. However, we have the additional constraint that $\Phi(x)$ is determined along a zero-energy trajectory. This reflects the fact that we are interested in a quasistationary density and thus should take $t_0 \to -\infty$ and x^* to be the fixed point x_-. The vanishing of the Lagrangian at the fixed points corresponds to a zero-energy solution.

10.3 Path-Integral Representation of a Birth–Death Master Equation

The connection between WKB methods for solving FPT problems and large-deviation variational principles also extends to chemical master equations in the large N limit (with N^{-1} playing the role of a weak noise parameter ε). Once again this connection can be established using path-integral methods, which were first developed for master equations and reaction–diffusion systems by Doi and Peliti [147, 148, 503]. For the sake of illustration, we will apply the "coherent states" formulation of Peliti to the birth–death master equation

$$\frac{dP(n,t)}{dt} = \omega_+(n-1)P(n-1,t) + \omega_+(n+1)P(n+1,t) - [\omega_+(n) + \omega_-(n)]P(n,t) \tag{10.3.1}$$

for $n = 0,\ldots,N$ with boundary conditions $P(-1,t) = P(N+1,t) = 0$. We are imagining an ensemble of N independent elements, each evolving according to a two-state Markov process with $\sigma_j = 0,1$ for $j = 1,\ldots,N$. Then $P(n,t)$ is the probability that there are n elements in the up-state ($\sigma = 1$) at time t. Such a system could describe N receptors or N ion channels, for example, with $\sigma = 1$ corresponding to a bound receptor or an open ion channel. We begin by introducing an abstract vector space (also known as a Fock space) with elements $|n\rangle$ representing the different occupancies of the state $\sigma = 1$. Introduce a pair of creation–annihilation linear operators that satisfy the commutation rule

$$[A, A^\dagger] \equiv AA^\dagger - A^\dagger A = 1. \tag{10.3.2}$$

These operators generate the full vector space by acting on the "vacuum" state $|0\rangle$, which represents the state in which all elements are in the down-state ($\sigma = 0$), with $A|0\rangle = 0$. The state $|n\rangle$ is then generated according to

$$|n\rangle = A^{\dagger n}|0\rangle. \tag{10.3.3}$$

Inner products in this state space are defined by $\langle 0|0\rangle = 1$ and the commutation relations. It follows that the dual of the vector $A^\dagger|0\rangle$ is $\langle 0|A$. The number of elements in the up-state can be extracted by operating on a state vector with the number operator $\Phi^\dagger\Phi$ and using the commutation relations:

$$A^\dagger A|n\rangle = n|n\rangle. \tag{10.3.4}$$

In order to see this, note that

$$\begin{aligned} A^\dagger A|n\rangle &= A^\dagger AA^\dagger|n-1\rangle = A^\dagger[A,A^\dagger]|n-1\rangle + A^\dagger AA^\dagger]|n-1\rangle \\ &= |n\rangle + A^\dagger A^\dagger A|n-1\rangle = 2|n\rangle + [A^\dagger]^3 A|n-2\rangle \\ &= n|n\rangle + [A^\dagger]^n A|0\rangle = n|n\rangle. \end{aligned}$$

Similarly, we have

$$A|n\rangle = n|n-1\rangle, \quad A^\dagger|n\rangle = |n+1\rangle. \tag{10.3.5}$$

The next step is to construct an operator representation of the master equation (10.3.1). Given the probability distribution $P(n,t)$, we introduce the state vector

$$|\phi(t)\rangle = \sum_n P(n,t)A^{\dagger n}|0\rangle = \sum_{n=0}^{N} P(n,t)|n\rangle. \tag{10.3.6}$$

Introducing the projection state

$$|\emptyset\rangle = \exp\left(A^\dagger\right)|0\rangle \tag{10.3.7}$$

with $A|\emptyset\rangle = |\emptyset\rangle$ and $\langle\emptyset|0\rangle = 1$, we can then express expectation values in terms of inner products. For example,

$$\langle\emptyset|A^\dagger A|\phi(t)\rangle = \sum_{n=0}^{N} nP(n,t) = \langle n(t)\rangle. \tag{10.3.8}$$

Differentiating the state vector $|\phi(t)\rangle$ with respect to t and using the master equation (10.3.1) gives

$$\begin{aligned}\partial_t|\phi(t)\rangle &= \sum_n \partial_t P(n,t)|n\rangle \\ &= \sum_n \left[\omega_+(n-1)P(n-1,t) - \omega_+(n)P(n,t)\right]|n\rangle \\ &\quad + \sum_n \left[\omega_-(n+1)P(n+1,t) - \omega_-(n)P(n,t)\right]|n\rangle \\ &= \left[(A - A^\dagger A)\overline{\omega}_-(B) + (A^\dagger - 1)\omega_+(B)\right]|\phi(t)\rangle, \end{aligned} \tag{10.3.9}$$

where $B = A^\dagger A$ and $\overline{\omega}_-(n) = \omega_-(n)/n$. We see that the operator representation of the master equation (10.3.1) is

$$\partial_t|\phi(t)\rangle = \mathscr{H}|\phi(t)\rangle \tag{10.3.10}$$

with

$$\mathscr{H}(A,A^\dagger) = (A - A^\dagger A)\overline{\omega}_-(B) + (A^\dagger - 1)\omega_+(B). \tag{10.3.11}$$

Note that the operator equation (10.3.10) can be converted to a PDE describing the evolution of the corresponding generating function $G(z,t)$ of the probability distribution $P(n,t)$. This follows from setting

$$G(z,t) \equiv \langle\emptyset|e^{zA}|\phi(t)\rangle = \sum_n z^n P(n,t)$$

and converting the operators $A, A^\dagger$ to $d/dz, z$. Both sets of operators satisfy the same commutation relations. Formally speaking, the solution to the operator version of the master equation (10.3.10) can be written as

$$|\phi(t)\rangle = e^{\mathscr{H}(A,A^\dagger)t}|\phi(0)\rangle, \tag{10.3.12}$$

and the expectation value of some physical quantity such as the number $n(t)$ can now be expressed as

$$\langle n(t)\rangle = \langle\emptyset|A^\dagger A e^{\mathscr{H}(A,A^\dagger)t}|\phi(0)\rangle. \tag{10.3.13}$$

Note that such an operator formalism, which is borrowed from quantum mechanics, has also been developed within the specific context of stochastic gene regulatory networks [558, 664].

In order to convert the operator form of the expectation value into a path integral, we divide the time interval $[0,t]$ into $\mathscr{N}$ intervals of length $\Delta t = t/\mathscr{N}$ and set $t_r = r\Delta t$, $r = 0,1,\ldots,\mathscr{N}$. We then introduce coherent states of the form

$$|\varphi(t)\rangle = \exp\left(-\frac{1}{2}|\varphi(t)|^2\right)\exp\left(\varphi(t)A^\dagger\right)|0\rangle, \tag{10.3.14}$$

such that $\varphi(t)$ is the complex-valued eigenvalue of the annihilation operator A, with complex conjugate φ^*. Coherent states satisfy the completeness relation

$$\int \frac{d\varphi d\varphi^*}{\pi}|\varphi\rangle\langle\varphi| = 1. \tag{10.3.15}$$

At each discrete-time step we insert a complete set of coherent states using the completeness relation (10.3.15) so that the expectation value becomes

$$\begin{aligned}\langle n(t)\rangle = {} & \langle\emptyset|A^\dagger A|\varphi(t)\rangle \\ & \left[\prod_{r=1}^{\mathscr{N}}\langle\varphi(t_r)|(1+\mathscr{H}\Delta t)|\varphi(t_{r-1})\rangle\right]\langle\varphi(0)|\phi(0)\rangle,\end{aligned} \tag{10.3.16}$$

where we have made use of the formula

$$\mathrm{e}^{\mathscr{H}t} = \lim_{\mathscr{N}\to\infty}(1+\mathscr{H}\Delta t)^{\mathscr{N}}.$$

The crucial observation is that at the rth time step we can replace the annihilation and creation operators $A, A^\dagger$ in $\mathscr{H}$ by the corresponding coherent state eigenvalues $\varphi^*(t_{r+1}), \varphi(t_r)$[2] Thus

$$\langle\varphi(t_r)|\varphi(t_{r-1})\rangle = (1+H(\varphi(t_{r-1}),\varphi^*(t_{r-1}))\Delta t)\langle\varphi(t_r)|\varphi(t_{r-1})\rangle,$$

where

$$H(\varphi,\varphi^*) = (\varphi-\varphi^*\varphi)\overline{\omega}_-(\psi)+(\varphi^*-1)\omega_+(\psi), \tag{10.3.17}$$

with $\psi = \varphi^*\varphi$ and

$$\begin{aligned}\langle\varphi(t_r)|\varphi(t_{r-1})\rangle &= \exp\left(-\frac{1}{2}|\varphi(t_r)|^2-\frac{1}{2}|\varphi(t_{r-1})|^2\right) \\ &\quad\times\langle 0|\exp\left(\varphi^*(t_r)A\right)\exp\left(\varphi(t_{r-1})A^\dagger\right)|0\rangle \\ &= \exp\left(-\frac{1}{2}|\varphi(t_r)|^2-\frac{1}{2}|\varphi(t_{r-1})|^2\right)\mathrm{e}^{\varphi^*(t_r)\varphi(t_{r-1})} \\ &= \exp\left(-\frac{1}{2}|\varphi(t_{r-1})|^2+\frac{1}{2}|\varphi(t_r)|^2\right)\mathrm{e}^{-\varphi^*(t_r)[\varphi(t_r)-\varphi(t_{r-1})]}.\end{aligned}$$

Combining a product of these terms for increasing r shows that the first exponential terms cancel except at the initial and final times, whereas the second exponential yields a factor

[2] Technically speaking, one first has to normal-order the operator $\mathscr{H}$ by moving all operators $A^\dagger$ to the right of all operators A using repeated application of the commutation rule. However, given the dependence of the transition rates on the system size N, this normal ordering introduces $O(1/N)$ corrections, which can be ignored to leading order.

$$\exp\left(-\varphi^*(t_r)\frac{d\varphi(t_r)}{dt}\Delta t + O(\Delta t^2)\right).$$

Moreover

$$\langle \emptyset | \varphi(t)\rangle = \mathrm{e}^{-|\varphi(t)|^2/2}\mathrm{e}^{\varphi(t)}.$$

Finally, taking the limits $\mathscr{N} \to \infty$ and $\Delta t \to 0$, we obtain the following path-integral representation of the expectation value:

$$\langle n(t)\rangle = \int D\varphi \int D\varphi^*\ \varphi(t)\mathrm{e}^{-S[\varphi,\varphi^*]}$$

where S is given by the action

$$S[\varphi,\varphi^*] = \int_0^t \varphi^*\left[\partial_\tau\varphi - H(\varphi,\varphi^*)\right]\, d\tau, \tag{10.3.18}$$

where we have dropped terms dependent on initial and final times. It turns out if the initial probability density is taken to be a Poisson distribution with mean $\bar{n}$, then the integration with respect to $\varphi(0)$ and $\varphi^*(0)$ simply enforces the initial condition $\varphi(0) = \bar{n}$. In anticipation of deriving a large deviation principle for large N, we now make explicit the N dependence of the transition rates,

$$\omega_\pm(n) = N\Omega_\pm(n/N).$$

It follows that $\overline{\omega}_-(n) = \overline{\Omega}_-(n/N)$ with $x\overline{\Omega}_-(x) = \Omega_-(x)$. In order to incorporate this scaling into the path integral, we rescale the variable φ according to $\varphi \to \varphi/N$. We then have for $x(t) = n(t)/N$

$$\langle x(t)\rangle = \int D\varphi \int D\varphi^*\ \varphi(t)\mathrm{e}^{-NS[\varphi,\varphi^*]}. \tag{10.3.19}$$

It follows from the saddle point method or steepest descents that in the large N limit, the path integral is dominated by the classical solution $\varphi = q, \varphi^* = p$, which is obtained by minimizing the action (10.3.18) with respect to time-dependent trajectories in the phase space $(\varphi(t), \varphi^*(t))$ with the initial condition $\varphi(0) = \bar{n}$ and final condition $\varphi(t) = \varphi$. Denoting the minimal action by $\Phi(\varphi, t|\bar{n})$, we see that

$$\langle n(t)\rangle \sim \int \varphi \mathrm{e}^{-N\Phi(\varphi,t|\bar{n})} d\varphi,$$

which has the form of a large deviation principle.

As in the case of the FP equation with weak noise, we have an effective Hamiltonian system in which the minimal action is evaluated along the trajectory (most probable path) given by the solution to Hamilton's equations

$$\dot{q} = \frac{\partial H}{\partial p} = \Omega_+(pq) - q\overline{\Omega}_-(pq) + (p-1)q\Omega'_+(pq) - (p-1)q^2\overline{\Omega}'_-(pq)$$
$$\dot{p} = -\frac{\partial H}{\partial q} = (p-1)\overline{\Omega}_-(pq) + (p-1)qp\overline{\Omega}'_-(pq) - (p-1)p\Omega'_+(pq),$$

where (after the N rescaling)

$$H(q,p) = (p-1)\Omega_+(pq) - (p-1)q\overline{\Omega}_-(pq).$$

We now make the observation that $p = 1$ is an invariant submanifold of the dynamics for which $\dot{p} = 0$ and

$$\dot{q} = \Omega_+(q) - q\overline{\Omega}_-(q) = \Omega_+(q) - \Omega_-(q),$$

which we recognize as the kinetic equation obtained in the deterministic limit $N \to \infty$. However, when $p \neq 1$, it is clear that we cannot identify q as the physical variable $n(t)/N$, since the transition rates depend on the product pq. (Moreover, in order to impose the initial condition $\varphi(0) = \bar{n}$ we had to take the initial distribution to be Poisson rather than $\delta_{n,\bar{n}}$.) Nevertheless, one can perform a canonical change of variables that allows us to identify the position variable as n/N, namely,

$$\tilde{q} = pq, \quad \tilde{p} = \ln(p). \tag{10.3.20}$$

The corresponding Hamiltonian becomes (after dropping tildes)

$$H(q,p) = (\mathrm{e}^p - 1)\Omega_+(q) + (\mathrm{e}^{-p} - 1)\Omega_-(q). \tag{10.3.21}$$

This is identical to the Hamiltonian (10.1.12) derived using WKB methods in Sect. 10.1.1, thus establishing the connection between WKB methods and variational principles for a birth–death master equation.

10.4 Path-Integral Representation of a Stochastic Hybrid System

Recall the differential Chapman–Kolmogorov (CK) equation (10.1.29) for a one-dimensional stochastic hybrid system with fast kinetics:

$$\frac{\partial p_n}{\partial t} = -\frac{\partial [F_n(x)p_n(x,t)]}{\partial x} + \frac{1}{\varepsilon}\sum_{n' \in I} A_{nn'}(x)p(x). \tag{10.4.1}$$

As highlighted in Sect. 10.1.2, the WKB analysis of this stochastic hybrid system differs from the WKB analysis of master equations (and FP equations) in the weak noise limit, since there is no obvious Hamiltonian structure, that is, the quasi-potential satisfies a matrix equation (10.1.42) rather than a scalar

Hamilton–Jacobi equation. Recently, a path-integral formulation of a stochastic hybrid system has been developed that provides a variational principle for the correct Hamiltonian [73]. We describe this construction for the one-dimensional case, although the extension to higher dimensions is straightforward. First, we discretize time by dividing a given interval $[0,T]$ into N equal subintervals of size Δt such that $T = N\Delta t$ and set $x_j = x(j\Delta t), n_j = n(j\Delta t)$. The conditional probability density for $x_1,\ldots,x_N$ given x_0 and a particular realization of the stochastic discrete variables $n_j, j = 0,\ldots,N-1$, is

$$P(x_1,\ldots,x_N|x_0,n_0,\ldots,n_{N-1}) = \prod_{j=0}^{N-1} \delta\left(x_{j+1} - x_j - F_{n_j}(x_j)\Delta t\right).$$

Inserting the Fourier representation of the Dirac delta function gives

$$\begin{aligned} P(x_1,\ldots,x_N|x_0,n_0,n_1,\ldots n_{N-1}) &= \prod_{j=0}^{N-1}\left[\int_{-\infty}^{\infty} \mathrm{e}^{-i\tilde{x}_j\left(x_{j+1}-x_j-F_{n_j}(x_j)\Delta t\right)}\frac{d\tilde{x}_j}{2\pi}\right] \\ &\equiv \prod_{j=0}^{N-1}\left[\int_{-\infty}^{\infty} H_{n_j}(x_{j+1},x_j,p_j)\frac{d\tilde{x}_j}{2\pi}\right]. \end{aligned}$$

On averaging with respect to the intermediate states $n_j, j = 1,N-1$, we have

$$P(x_1,\ldots,x_N|x_0,n_0) = \left[\prod_{j=0}^{N-1}\int_{-\infty}^{\infty}\frac{d\tilde{x}_j}{2\pi}\right]\sum_{n_1,\ldots,n_{N-1}}\prod_{j=0}^{N-1} T_{n_{j+1},n_j}(x_j)H_{n_j}(x_{j+1},x_j,p_j)$$

where

$$\begin{aligned} T_{n_{j+1},n_j}(x_j) &\sim A_{n_{j+1},n_j}(x_j)\frac{\Delta t}{\varepsilon} + \delta_{n_{j+1},n_j}\left(1 - \sum_m A_{m,n_j}(x_j)\frac{\Delta t}{\varepsilon}\right) + o(\Delta t) \\ &= \left(\delta_{n_{j+1},n_j} + A_{n_{j+1},n_j}(x_j)\frac{\Delta t}{\varepsilon}\right). \end{aligned}$$

Consider the eigenvalue equation

$$\sum_m \left[A_{nm}(x) + q\delta_{n,m}F_m(x)\right]R_m^{(s)}(x,q) = \Lambda_s(x,q)R_n^{(s)}(x,q), \tag{10.4.2}$$

and let $\xi_m^{(s)}$ be the adjoint eigenvector. Inserting multiple copies of the identity

$$\sum_s \xi_m^{(s)}(x,q)R_n^{(s)}(x,q) = \delta_{m,n},$$

with $(x,q) = (x_j,q_j)$ at the jth time step, we have

$$
\begin{aligned}
&T_{n_{j+1}n_j}(x_j)H_{n_j}(x_{j+1},x_j,p_j)\\
&\sim \sum_{s_j,m} R^{(s_j)}_{n_{j+1}}(x_j,q_j)\xi^{(s_j)}_m(x_j,q_j)\left(\delta_{n_j,m}+A_{mn_j}(x_j)\frac{\Delta t}{\varepsilon}\right)H_{n_j}(x_{j+1},x_j,p_j)\\
&=\sum_{s_j}\left(1+[\Lambda_{s_j}(x_j,q_j)-q_jF_{n_j}(x_j)]\frac{\Delta t}{\varepsilon}\right)\mathrm{e}^{-i\tilde{x}_j\left(x_{j+1}-x_j-F_{n_j}(x_j)\Delta t\right)}\\
&\qquad\times R^{(s_j)}_{n_{j+1}}(x_j,q_j)\xi^{(s_j)}_{n_j}(x_j,q_j)\\
&\sim\sum_{s_j}\exp\left([\Lambda_{s_j}(x_j,q_j)-q_jF_{n_j}(x_j)]\frac{\Delta t}{\varepsilon}-i\tilde{x}_j\left(x_{j+1}-x_j-F_{n_j}(x_j)\Delta t\right)\right)\\
&\qquad\times R^{(s_j)}_{n_{j+1}}(x_j,q_j)\xi^{(s_j)}_{n_j}(x_j,q_j)\\
&=\sum_{s_j}\exp\left(\left[\Lambda_{s_j}(x_j,q_j)-q_j\frac{x_{j+1}-x_j}{\Delta t}\right]\frac{\Delta t}{\varepsilon}\right)\exp\left([i\varepsilon p_jF_{n_j}(x_j)-q_jF_{n_j}(x_j)]\frac{\Delta t}{\varepsilon}\right)\\
&\qquad\times R^{(s_j)}_{n_{j+1}}(x_j,q_j)\xi^{(s_j)}_{n_j}(x_j,q_j),
\end{aligned}
$$

to leading order in $O(\Delta x,\Delta t)$. Substituting into the expression for P and integrating over intermediate states x_j lead to

$$
\begin{aligned}
P(x_N,n_N|x_0,n_0)&=\prod_{j=1}^{N-1}\int_{-\infty}^{\infty}dx_jP(x_1,\ldots,x_N,n_N|x_0,n_0) \qquad (10.4.3)\\
&\left[\prod_{j=1}^{N-1}\int_{-\infty}^{\infty}dx_j\right]\left[\prod_{j=0}^{N-1}\int_{-\infty}^{\infty}\frac{dp_j}{2\pi}\right]\sum_{n_1,\ldots,n_{N-1}}\sum_{s_0,\ldots,s_{N-1}}\left[\prod_{j=0}^{N-1}R^{(s_j)}_{n_{j+1}}(x_j,q_j)\xi^{(s_j)}_{n_j}(x_j,q_j)\right]\\
&\qquad\exp\left(\sum_j\left[\Lambda_{s_j}(x_j,q_j)-q_j\frac{x_{j+1}-x_j}{\Delta t}\right]\frac{\Delta t}{\varepsilon}\right)\exp\left([i\varepsilon p_jF_{n_j}(x_j)-q_jF_{n_j}(x_j)]\frac{\Delta t}{\varepsilon}\right).
\end{aligned}
$$

By inserting the eigenfunction products and using the Fourier representation of the Dirac delta function, we have introduced sums over the discrete labels s_j and new phase variables p_j. However, this representation allows us to derive a large deviation principle in the limit $\varepsilon\to 0$. First, note that the discretized path integral is independent of the q_j. Therefore, we are free to set $q_j=i\varepsilon p_j$ for all j, thus eliminating the final exponential factor. This choice means that we can perform the summations with respect to the intermediate discrete states n_j using the orthogonality relation

$$
\sum_n R^{(s)}_n(x_j,q_{j-1})\xi^{(s')}_n(x_{j+1},q_j)=\delta_{s,s'}+O(\Delta x,\Delta q).
$$

We thus obtain the result that $s_j=s$ for all j, which means that we can then take the continuum limit of equation (10.4.3) to obtain the following path integral from $x(0)=x_0$ to $x(\tau)=x$ (after performing the change of variables $i\varepsilon p_j\to p_j$ (complex contour deformation)):

$$P(x,n,\tau|x_0,n_0,0) \tag{10.4.4}$$
$$= \sum_s \int_{x(0)=x_0}^{x(\tau)=x} \exp\left(-\frac{1}{\varepsilon}\int_0^\tau [p\dot{x} - \Lambda_s(x,p)]dt\right) R_n^{(s)}(x,p(\tau))\xi_{n_0}^{(s)}(x_0,p(0))\,\mathscr{D}[p]\mathscr{D}[x].$$

Applying the Perron–Frobenius theorem to the linear operator on the left-hand side of equation (10.4.2) shows that there exists a real, simple Perron eigenvalue labeled by $s=0$, say, such that $\Lambda_0 > \mathrm{Re}(\Lambda_s)$ for all $s>0$. It follows that in the limit $\varepsilon \to 0$, the largest contributions to the path integral (10.4.4) and the most likely paths in phase space (x,p) are obtained by restricting the sum over s to $s=0$. Also note that the factor $R_n^{(0)}(x,p(\tau))\xi_{n_0}^{(0)}(x_0,p(0))$ in equation (10.4.4) essentially projects on to stochastic trajectories that start in the discrete state n_0 and terminate in the discrete state n. We will ignore any restrictions on these discrete states and simply consider the probability density (for fixed $x(0)=x_0$)

$$P(x,t) = \int_{\mathbf{x}(0)=\mathbf{x}_0}^{\mathbf{x}(\tau)=\mathbf{x}} D[x]D[p]\mathrm{e}^{-S[x,p]/\varepsilon}, \tag{10.4.5}$$

with the action

$$S[x,p] = \int_0^\tau [p\dot{x} - \Lambda_0(x,p)]\,dt. \tag{10.4.6}$$

We now have a classical variational problem, in which the Perron eigenvalue $\Lambda_0(x,p)$ is identified as a Hamiltonian and the most probable path is the solution to Hamilton's equations

$$\dot{x} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial x}, \quad H(x,p) = \Lambda_0(x,p). \tag{10.4.7}$$

Note that the same action principle can also be derived using large deviation theory, as detailed in the monograph by Kifer [332].

Now suppose that we have a higher-dimensional stochastic hybrid system with M continuous variables x_α, $\alpha = 1,\ldots M$, and a single discrete variable $n=0,\ldots,K-1$. The multivariate CK equation takes the form

$$\frac{\partial p}{\partial t} = -\sum_{\alpha=1}^M \frac{\partial}{\partial x_\alpha}(F_\alpha(\mathbf{x},n)p(\mathbf{x},n,t)) + \frac{1}{\varepsilon}\sum_m A(n,m;\mathbf{x})p(\mathbf{x},m,t). \tag{10.4.8}$$

The drift terms $F_\alpha(\mathbf{x},n)$ for fixed n represent the piecewise deterministic dynamics according to

$$\tau\frac{du_\alpha}{dt} = F_\alpha(\mathbf{x},n), \quad \alpha = 1,\ldots,M. \tag{10.4.9}$$

Following along identical lines to the one-dimensional case, we can derive a path-integral representation of the solution to equation (10.4.8):

$$p(\mathbf{x},n,\tau|\mathbf{x}_0,n_0,0) = \int_{\mathbf{x}(0)=\mathbf{x}_0}^{\mathbf{x}(\tau)=\mathbf{x}} \mathscr{D}[\mathbf{p}]\mathscr{D}[\mathbf{x}] \exp\left(-\frac{1}{\varepsilon}S[\mathbf{x},\mathbf{p}]\right) R^{(0)}(\mathbf{x},\mathbf{p}(\tau),n)\xi^{(0)}(\mathbf{x}_0,\mathbf{p}(0),n_0) \quad (10.4.10)$$

with action

$$S[\mathbf{x},\mathbf{p}] = \int_0^\tau \left[\sum_{\alpha=1}^M p_\alpha \dot{x}_\alpha - \Lambda_0(\mathbf{x},\mathbf{p})\right] dt. \quad (10.4.11)$$

Here Λ_0 is the Perron eigenvalue of the following linear operator equation (cf. equation (10.4.2)):

$$\sum_m A(n,m;\mathbf{x})R^{(0)}(\mathbf{x},\mathbf{p},m) = [\Lambda_0(\mathbf{x},\mathbf{p}) - \sum_{\alpha=1}^M p_\alpha F_\alpha(\mathbf{x},n)]R^{(0)}(\mathbf{x},\mathbf{p},n),$$

and $\xi^{(0)}$ is the adjoint eigenvector. Suppose that the underlying deterministic system (10.4.9) has a unique stable fixed point $\mathbf{x}_*$. The quasi-potential of the corresponding stationary density can then be obtained by finding zero-energy solutions of Hamilton's equations

$$\dot{\mathbf{x}} = \nabla_{\mathbf{p}}\mathscr{H}(\mathbf{x},\mathbf{p}), \quad \dot{\mathbf{p}} = -\nabla_{\mathbf{x}}\mathscr{H}(\mathbf{x},\mathbf{p}), \quad (10.4.12)$$

with $\mathbf{x} = (x,y), \mathbf{p} = (p_x,p_y)$. If such a solution can be found, then we can construct a quasi-potential Φ by identifying it as the action along a zero-energy solution curve $\mathbf{x}(t)$. That is,

$$\frac{d\Phi}{dt} \equiv \sum_{\alpha=1}^M \frac{\partial\Phi}{\partial x_\alpha}\frac{dx_\alpha}{dt} = \sum_{\alpha=1}^M p_\alpha \frac{dx_\alpha}{dt}, \quad (10.4.13)$$

with $p_\alpha = \partial\Phi/\partial x_\alpha$.

10.5 Excitability in the Stochastic Morris–Lecar Model

In this final section, we apply WKB methods to the stochastic version of the Morris–Lecar model introduced in Sect. 3.5. The deterministic model is given by the planar dynamical system

$$\frac{dx}{dt} = a(x)f_{\mathrm{Na}}(x) + wf_{\mathrm{K}}(x) + f_L(x), \quad (10.5.1a)$$

$$\frac{dw}{dt} = \frac{w_\infty(x) - w}{\tau_w(x)}, \quad (10.5.1b)$$

where the membrane voltage is denoted by x rather than V, and

$$\tau_w(x) = \frac{1}{\alpha_w(x) + \beta_w(x)}, \qquad w_\infty(x) = \alpha_w(x)\tau_w(x).$$

As we showed in Sect. 3.5, the generation of action potentials in the deterministic model can be analyzed using a slow/fast analysis, since the dynamics of the recovery variable w (representing the fraction of open K ion channels) is slow relative to that of the membrane voltage x; see Fig. 3.14. In the analysis of membrane voltage fluctuations, it was assumed that the potassium channel dynamics could be ignored during initiation of a SAP. This corresponds to keeping the recovery variable w fixed. The resulting stochastic bistable model supported the generation of SAPs due to fluctuations in the opening and closing of fast Na channels. However, it turns out that this slow/fast analysis breaks down when the effects of K channel noise are included [476]. That is, it is possible to generate a SAP due to fluctuations causing several K channels to close simultaneously, effectively decreasing w and thereby causing v to rise. This can be confirmed by numerically solving the full stochastic model as illustrated in Fig. 10.3. It follows that keeping w fixed in the stochastic model excludes the latter mechanism, and thus the resulting MFPT calculation underestimates the

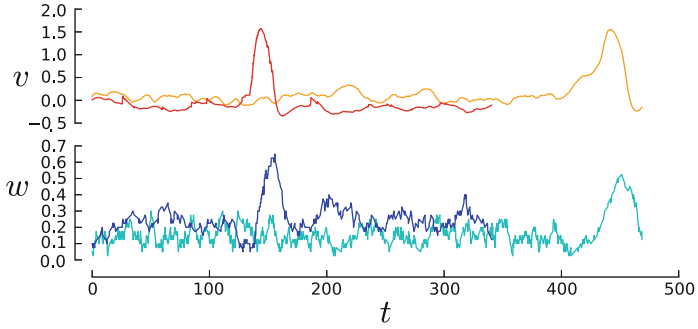


Fig. 10.3: Noise-induced SAPs. Stochastic Na and K channels: *red* for $v(t)$ and *blue* for $w(t)$. Stochastic K and deterministic Na: *orange* for voltage $v(t)$ and *light blue* for $w(t)$. Parameter values can be found in [476]

spontaneous rate of action potentials. In order to investigate the above phenomenon, it is necessary to consider the full stochastic ML model given by equations (3.5.13) and (3.5.14) with N sodium channels and M potassium channels. An additional complication is that the full model is an excitable rather than a bistable system, so it is not straightforward to relate the generation of SAPs with a noise-induced escape problem. Nevertheless, Newby et al. [476] used WKB methods to identify the most probable paths of escape from the resting state and obtained the following results, which are illustrated in Fig. 10.4:

(i) Most probable paths of escape dip significantly below the resting value for w, indicating a breakdown of the deterministic slow/fast decomposition.

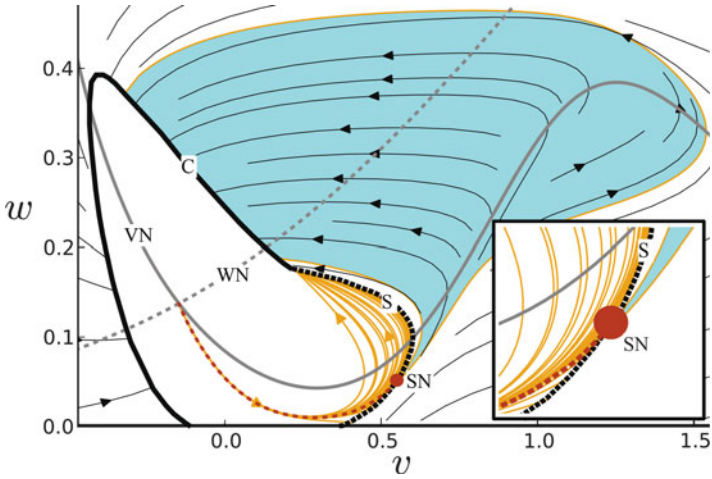


Fig. 10.4: Most probable paths of escape from the resting state of the stochastic ML model calculated using a WKB approximation [476]. All paths of escape that enter the *shaded blue region* represent large excursions in state space and coincide with SAPs. All of the SAP trajectories are initially bunched together (*red dashed* curve) until they cross the bottleneck or metastable saddle node (SN). Curves that don't pass through SN are bounded by a curve (S) that acts like a stochastic separatrix. Also shown are a caustic (C) where paths of escape intersect, the v nullcline (VN), and the w nullcline (WN). The resting state is surrounded by an effective "basin of attraction" bounded by C and S. Here $N = M = 40$ and $\varepsilon = 0.1$. Other parameter values can be found in [476]

(ii) Escape trajectories all pass through a narrow region of state space (bottleneck or stochastic saddle node) so that although there is no well-defined separatrix for an excitable system, it is possible to formulate an escape problem by determining the MFPT to reach the bottleneck from the resting state.

As we have already made clear, there is no a priori reason to identify the WKB Hamiltonian as the "correct " Hamiltonian, in the sense that solutions of Hamilton's equations represent maximum likelihood paths of the underlying stochastic system. Indeed, one finds that the Perron eigenvalue obtained using the path-integral formulation [73] or large deviation theory [332] actually differs from the WKB Hamiltonian, but the two are equivalent on zero-energy surfaces so that the WKB method is still valid. We now develop these results in more detail. Let $p(x,m,n,t)$ denote the probability density for the stochastic ML model, which evolves according to the differential Chapman–Kolmogorov (CK) equation,

$$\frac{\partial p}{\partial t} = -\frac{\partial (Fp)}{\partial x} + \mathbb{L}_{\mathrm{K}} p + \mathbb{L}_{\mathrm{Na}} p, \tag{10.5.2}$$

with $F = F(x,m,n)$ given by equation (3.5.13):

$$F(x,m,n) = \frac{n}{N} f_{\mathrm{Na}}(x) + \frac{m}{M} f_{\mathrm{K}}(x) - g(x). \tag{10.5.3}$$

The jump operators are defined according to (dropping the explicit t dependence)

$$\begin{aligned}\mathbb{L}_{\rm Na}p(x,m,n) &= \omega^+_{\rm Na}(n+1,x)p(x,m,n+1)\\ &\quad+\omega^-_{\rm Na}(n-1,x)p(x,m,n-1)-[\omega^+_{\rm Na}(n,x)+\omega^+_{\rm Na}(n,x)]p(x,m,n)\end{aligned} \tag{10.5.4a}$$

$$\begin{aligned}\mathbb{L}_{\rm K}p(x,m,n) &= \omega^+_{\rm K}(m+1,x)p(x,m+1,n)+\omega^-_{\rm K}(m-1,x)p(x,m-1,n)\\ &\quad-[\omega^+_{\rm K}(m,x)+\omega^+_{\rm K}(m,x)]p(x,m,n)\end{aligned} \tag{10.5.4b}$$

with

$$\omega^+_{\rm Na}(n,x)=n\beta_{\rm Na}(x),\quad \omega^-_{\rm Na}(n,x)=(N-n)\alpha_{\rm Na}(x), \tag{10.5.5a}$$

$$\omega^+_{\rm K}(m,x)=m\beta_{\rm K}(x),\quad \omega^-_{\rm K}(m,x)=(M-m)\alpha_{\rm K}(x). \tag{10.5.5b}$$

In contrast to the bistable sodium ion channel model, we cannot treat both the Na and K channel kinetics as fast, and therefore we cannot develop a variational problem by scaling all transition rates in terms of a small parameter ε and applying the analysis of Sect. 10.1.3. In fact, rather than a piecewise deterministic system, we now have a multi-scale stochastic system, in which both fast and slow processes are intrinsically stochastic. Multi-scale stochastic processes also arise in models of gene regulatory networks [470]. Instead, we treat $n(t)$ as a fast variable by taking $\alpha_{\rm Na},\beta_{\rm Na}=\mathscr{O}(1/\varepsilon)$, whereas we treat $w(t)=m(t)/M$ as a continuous (recovery) variable with $M=1/\varepsilon$. Setting $p=p(x,w,n,t)$ and rescaling the transition rates for sodium by a factor of $1/\varepsilon$, the CK equation (10.5.2) for p becomes

$$\frac{\partial p}{\partial t}=-\frac{\partial(Fp)}{\partial x}+\frac{1}{\varepsilon}\mathbb{L}_{\rm K}p+\frac{1}{\varepsilon}\mathbb{L}_{\rm Na}p, \tag{10.5.6}$$

with jump operators

$$\begin{aligned}\mathbb{L}_{\rm Na}p(x,w,n) &= \Omega^+_{\rm Na}(n+1,x)p(x,w,n+1)+\Omega^-_{\rm Na}(n-1,x)p(x,m,n-1)\\ &\quad-[\Omega^+_{\rm Na}(n,x)+\Omega^+_{\rm Na}(n,x)]p(x,m,n)\end{aligned} \tag{10.5.7a}$$

$$\begin{aligned}\mathbb{L}_{\rm K}p(x,w,n) &= \Omega^+_{\rm K}(w+\varepsilon,x)p(x,w+\varepsilon,n)+\Omega^-_{\rm K}(w-\varepsilon,x)p(x,w-\varepsilon,n)\\ &\quad-[\Omega^+_{\rm K}(w,x)+\Omega^+_{\rm K}(w,x)]p(x,w,n),\end{aligned} \tag{10.5.7b}$$

and

$$\varepsilon^{-1}\Omega^\pm_{\rm Na}(n,x)=\omega^\pm_{\rm Na}(n,x),\quad \varepsilon^{-1}\Omega^\pm_{\rm K}(w,x)=\omega^\pm_{\rm Na}(w/\varepsilon,x).$$

In the limit $\varepsilon\to 0$, we recover the standard deterministic ML model (10.5.1).

Newby et al. [476] considered a WKB solution of the stationary CK equation (10.5.6) of the form

$$p=\phi_\varepsilon(x,w,n)=R_n(x,w)\mathrm{e}^{-\Phi(x,w)/\varepsilon}, \tag{10.5.8}$$

After substituting (10.5.8) into (10.5.6) (with $\partial p/\partial t=0$) and collecting terms in ε, one finds to leading order that

$$[\mathbb{L}_{\rm Na}+F_n(x,w)p_x+h(x,w,p_w)]R_n(x,w)=0, \tag{10.5.9}$$

where $p_x = \partial_x \Phi$, $p_w = \partial_w \Phi$,

$$h(x,w,p_w) = \sum_{j=\pm 1} (e^{-jp_w} - 1)\Omega_{\mathrm{K}}^{\pm}(w,x), \tag{10.5.10}$$

and $F_n(x,w) = F(x,Mw,n)$; see equation (3.5.13). Equation (10.5.9) can be solved for Φ and R_n using the ansatz $R_n(x,w) = A(x,w)^n/(n!(N-n)!)$. The function $A(x,w)$ is determined by substituting R_n into (10.5.9) to obtain a self-consistency condition with two terms: one linear in n and one independent of n. From the former we obtain

$$A(x,w) = \alpha_{\mathrm{Na}}(x) - \frac{1}{N}(p_x g(x,w) + h(x,w,p_w)), \tag{10.5.11}$$

where $g(x,w) = w f_{\mathrm{K}}(x) + f_L(x)$. After substituting (10.5.11) into the remaining n-independent term, we obtain a nonlinear scalar PDE for Φ in the form of a Hamilton–Jacobi equation [476]:

$$H_0(x,w,\partial_x \Phi, \partial_w \Phi) = 0, \tag{10.5.12}$$

with

$$\begin{aligned} H_0(x,w,p_x,p_w) &= (2g(x,w) + f_{\mathrm{Na}}(x))p_x h(x,w,p_w) \\ &+ (f_{\mathrm{Na}}(x) + g(x,w))g(x,w)p_x^2 + h(x,w,p_w)^2 \\ &- \frac{N}{1 - w_\infty(x)}\left([w_\infty(x) f_{\mathrm{Na}}(x) + g(x,w)]p_x + h(x,w,p_w)\right). \end{aligned} \tag{10.5.13}$$

Newby et al. [476] used the WKB Hamiltonian H_0 to generate maximum-likelihood paths such as those illustrated in Fig. 10.4. Moreover, they showed that conclusions based on WKB appeared to be consistent with Monte Carlo simulations of the full system (3.5.13). A more rigorous justification of these results can be obtained by constructing a path-integral representation. Here we simply write down the action and determine the Perron eigenvalue. The action takes the form

$$S[x,w,p_x,p_w] = \int_0^\tau [p_x \dot{x} + p_w \dot{w} - \Lambda_0(x,w,p_x,p_w)]dt \tag{10.5.14}$$

where Λ_0 is the Perron eigenvalue of the following linear operator equation (cf. equation (10.4.2))

$$\Lambda_0 R_n^{(0)} = [\mathbb{L}_{\mathrm{Na}} + F_n(x,w)p_x + h(x,w,p_w)]\, R_n^{(0)}. \tag{10.5.15}$$

Maximum likelihood paths correspond to extreme of the action in the full phase space $(\mathbf{x},\mathbf{p})$, where $\mathbf{x} = (x,w), \mathbf{p} = (p_x,p_w)$. These are thus solutions of Hamilton's equation with Λ_0 identified as the Hamiltonian. Equation (10.5.15) can be solved along similar lines to (10.5.9) using the ansatz $R_n(x,w) = A(x,w)^n/(n!(N-n)!)$. Collecting terms linear in n gives

$$A(x,w) = \alpha_{\mathrm{Na}}(x) - \frac{1}{N}(p_x g(x,w) + h(x,w,p_w) - \Lambda_0(x,w,p_x,p_w)),$$

whereas collecting terms independent of n and substituting for $A(x,w)$ gives the following quadratic equation for Λ_0:

$$\Lambda_0^2 - (2h(x,w,p_w) + \sigma(x,w,p_x))\Lambda_0 + H_0(x,w,p_x,p_w) = 0, \tag{10.5.16}$$

with

$$\sigma(x,w,p_x) = (2g(x) + f(x))p_x - N/(1 - w_\infty(x))$$

and H_0 the WKB Hamiltonian (10.5.13).

We have thus established that the "correct" Hamiltonian Λ_0 underlying maximum likelihood (optimal) paths of the stochastic Morris–Lecar model differs from the WKB Hamiltonian H_0. However, it turns out that both Hamiltonians generate the same optimal paths on the zero-energy surface, and thus the results of Newby et al. [476] still hold. First note that $\Lambda_0 = 0$ implies that $H_0 = 0$. Second, optimal paths are given by solutions of Hamilton's equations

$$\dot{\mathbf{x}} = \nabla_{\mathbf{p}} \Lambda_0(\mathbf{x},\mathbf{p}), \quad \dot{\mathbf{p}} = -\nabla_{\mathbf{x}} \Lambda_0(\mathbf{x},\mathbf{p}). \tag{10.5.17}$$

Given a solution curve $(x(t), w(t))$, known as a ray or metastable trajectory, the quasi-potential Φ can be determined along the ray by solving the equation

$$\frac{d\Phi}{dt} \equiv \frac{\partial \Phi}{\partial x}\frac{dx}{dt} + \frac{\partial \Phi}{\partial w}\frac{dw}{dt} = p_x \frac{dx}{dt} + p_w \frac{dw}{dt}. \tag{10.5.18}$$

Thus, the quasi-potential Φ can be identified as the action along a zero-energy trajectory. In order to show that the optimal paths also correspond to solutions of Hamilton's equations for the WKB Hamiltonian H_0, we differentiate the quadratic equation with respect to each of the variables $z = x, w, p_x, p_w$ and then set $\Lambda_0 = 0$. This shows that

$$\frac{\partial H_0}{\partial z} - [2h(x,w,p_w) + \sigma(x,w,p_x)]\frac{\partial \Lambda_0}{\partial z} = 0$$

and hence

$$\dot{x} = \frac{\partial \Lambda_0}{\partial p_x} = [2h(x,w,p_w) + \sigma(x,w,p_x)]^{-1}\frac{\partial H_0}{\partial p_x},$$

$$\dot{w} = \frac{\partial \Lambda_0}{\partial p_w} = [2h(x,w,p_w) + \sigma(x,w,p_x)]^{-1}\frac{\partial H_0}{\partial p_w},$$

$$\dot{p_x} = -\frac{\partial \Lambda_0}{\partial x} = -[2h(x,w,p_w) + \sigma(x,w,p_x)]^{-1}\frac{\partial H_0}{\partial x},$$

$$\dot{p_y} = -\frac{\partial \Lambda_0}{\partial w} = -[2h(x,w,p_w) + \sigma(x,w,p_x)]^{-1}\frac{\partial H_0}{\partial w}.$$

Since all four equations have the same scale factor, we see that the phase-space trajectories on the zero-energy surface of Λ_0 are identical to those on the zero energy surface of H_0.

10.6 Exercises

Problem 10.1 (WKB analysis of a birth–death process). Consider the birth–death master equation

$$\frac{d}{dt}P(n,t) = \omega_+(n-1)P(n-1,t) + \omega_-(n+1)P(n+1,t) - [\omega_+(n) + \omega_-(n)]P(n,t),$$

with transition rates of the form

$$\omega_+(n) = Nf(n/N), \quad \omega_-(x) = N.$$

Here f is the nonlinear sigmoid function

$$f(x) = \frac{f_0}{1 + e^{-\gamma(x-\theta)}}$$

for constants f_0, γ, and θ. The quasi-potential Φ of the associated quasistationary solution satisfies the Hamilton–Jacobi equation (see (10.1.12)), $H(x, \Phi'(x)) = 0$, where

$$H(x,p) = \Omega_+(x)[e^p - 1] + \Omega_-(x)[e^{-p} - 1],$$

and $\omega_\pm(Nx) = N\Omega_\pm(x)$.

(a) Determine Hamilton's equations for the given transition rates.
(b) Construct a phase portrait in the (x,p)-plane of solutions to Hamilton's equations (analogous to Fig. 10.2) for the parameter values $f_0 = 2$, $\gamma = 4$ and $\theta = 1.0$. Identify the fixed points of the corresponding deterministic system and the various zero-energy solutions. Distinguish between the activation and relaxation trajectories.

Problem 10.2 (WKB analysis of a conductance-based model with two types of fast ion channels). This problem extends the analysis of the stochastic ion channel model presented in Sect. 10.1.3. Consider a conductance-based model of a neuron that consists of two distinct, independent types of ion channel labeled by x and y, respectively. Let N be the total number of each type. Suppose that at time t, there are $n_j(t)$ open ion channels of type j with the remaining $N - n_j(t)$ channels closed. The membrane voltage v then evolves according to the Hodgkin–Huxley-type model

$$C\frac{dv}{dt} = \bar{g}_x \frac{n_x(t)}{N}(V_x - v) + \bar{g}_y \frac{n_y(t)}{N}(V_y - v),$$

where V_x, V_y are reversal potentials, $\bar{g}_x, \bar{g}_y$ are conductances, and C is a membrane capacitance. (Neglect any leak currents and set $C = 1$.) The voltage equation only holds between jumps in the discrete random variables n_x, n_y, which are given by the birth–death processes

$$n_j \underset{\omega_+^j(n_j)/\varepsilon}{\to} n_j + 1, \quad n_j \underset{\omega_-^j(n_j)/\varepsilon}{\to} n_j - 1,$$

for $j = x, y$, where the small parameter ε reflects the fast kinetics of the ion channels. The rescaled transition rates are

$$\omega_+^j(n_j) = \alpha_j(V)(N - n_j), \quad \omega_-^j(n_j) = \beta_j n_j.$$

(a) Write down the Chapman–Kolmogorov equation for the joint probability density p,

$$p(v, n_x, n_y, t)dv = \mathbb{P}[v \le V(t) \le v + dv, n_x(t) = n_x, n_y(t) = n_y],$$

given some initial condition $v(0) = v_0, n_j(0) = \bar{n}_j$.

(b) Derive the deterministic conductance-based model in the limit $\varepsilon \to 0$.

(c) Seek a quasistationary solution of the CK equation of the WKB form

$$\varphi(v, \mathbf{n}) = R(v, \mathbf{n}) \exp\left(-\frac{\Phi(v)}{\varepsilon}\right),$$

where $\Phi(v)$ is the quasi-potential. Introducing the asymptotic expansions $R \sim R^{(0)} + \varepsilon R^{(1)}$ and $\Phi \sim \Phi_0 + \varepsilon \Phi_1$, derive the leading order matrix equation for $R^{(0)}$ and Φ_0, which is the generalization of equation (10.1.42). Find the two positive solutions of this matrix equation for $\Phi_0'(v)$ using the ansatz

$$R^{(0)}(v, \mathbf{n}) = \frac{1}{[1 + \Gamma_x(v)]^N} \frac{1}{[1 + \Gamma_y(v)]^N} \frac{N![\Gamma_x(v)]^{n_x}}{(N - n_x)! n_x!} \cdot \frac{N![\Gamma_y(v)]^{n_y}}{(N - n_y)! n_y!}.$$

One of these solutions is $\Phi_0' = 0$, with $R^{(0)}(v, \mathbf{n})$ equal to the stationary solution of the pure jump Markov process, whereas the other is the desired WKB solution.

Problem 10.3 (Stochastic Morris–Lecar model). Consider the Chapman–Kolmogorov equation (10.5.6) of the stochastic Morris–Lecar model.

(a) Substitute the WKB solution

$$p = \phi_\varepsilon(x, w, n) = R_n(x, w) \mathrm{e}^{-\Phi(x,w)/\varepsilon},$$

into the stationary version of (10.5.6). Collecting terms in ε, show that to leading order

$$[\mathbb{L}_{\mathrm{Na}} + F_n(x, w) p_x + h(x, w, p_w)] R_n(x, w) = 0,$$

where $p_x = \partial_x \Phi$, $p_w = \partial_w \Phi$,

$$h(x,w,p_w) = \sum_{j=\pm 1} (e^{-jp_w} - 1)\Omega_{\mathrm{K}}^{\pm}(w,x),$$

$F_n(x,w) = F(x,Mw,n)$, and the operator $\mathbb{L}_{\mathrm{Na}}$ is given by equation (10.5.7a).

(b) Substituting the ansatz $R_n(x,w) = A(x,w)^n/(n!(N-n)!)$ and collecting terms linear in n and independent of n, respectively, show that

$$A(x,w) = \alpha_{\mathrm{Na}}(x) - \frac{1}{N}(p_x g(x,w) + h(x,w,p_w)),$$

where $g(x,w) = w f_{\mathrm{K}}(x) + f_L(x)$, and Φ satisfies the Hamilton–Jacobi equation

$$H_0(x,w,\partial_x \Phi, \partial_w \Phi) = 0,$$

with

$$\begin{aligned} H_0(x,w,p_x,p_w) &= (2g(x,w) + f_{\mathrm{Na}}(x))p_x h(x,w,p_w) \\ &+ (f_{\mathrm{Na}}(x) + g(x,w))g(x,w)p_x^2 + h(x,w,p_w)^2 \\ &- \frac{N}{1 - w_\infty(x)}\left([w_\infty(x) f_{\mathrm{Na}}(x) + g(x,w)]p_x + h(x,w,p_w)\right). \end{aligned}$$

Problem 10.4 (Perron eigenvalue of a binary stochastic hybrid system). Suppose that a continuous variable evolves according to piecewise dynamics on some finite interval (a,b),

$$\dot{x} = F_n(x), \quad n = 0,1, \tag{10.6.1}$$

with F_0, F_1 continuous and locally Lipschitz. Suppose that F_0, F_1 are non-vanishing within the interval (a,b), and $F_n(a) \geq 0, F_n(b) \leq 0$ for $n = 0,1$; the dynamics is then confined to (a,b). Denote the transition rates of the two-state Markov chain by $\omega_{\pm}(x)$ with

$$\{n=0\} \underset{\omega_-(x)}{\overset{\omega_+(x)}{\rightleftharpoons}} \{n=1\}.$$

(a) Write down the matrix $\mathbf{A}(x)$ with

$$A_{nm} = W_{nm} - \delta_{nm} \sum_p W_{pm}.$$

(b) Express the Perron eigenvalue equation

$$\sum_m [A_{nm}(x) + p\delta_{n,m} F_m(x)]\, R_m^{(0)}(x,p) = \Lambda_0(x,p) R_n^{(0)}(x,p),$$

as a 2×2 matrix equation

(c) Solve the resulting characteristic equation for the Perron eigenvalue Λ_0 and show that Λ_0 is real.

(d) Using Hamilton's equation $\dot{x} = \partial \Lambda_0 / \partial p$, derive the equation

$$\dot{x} = \frac{F_0(x) + F_1(x)}{2} + \frac{F_0(x) - F_1(x)}{2} \frac{p(F_0 - F_1) - (\omega_+ - \omega_-)}{\sqrt{[p(F_0 - F_1) - (\omega_+ - \omega_-)]^2 + \omega_+ \omega_-}}$$

and show that this can be rewritten in the form

$$\dot{x} = F_0(x)\psi_0(x) + F_1(x)\psi_1(x),$$

with $\psi_{0,1}$ non-negative functions and $\psi_0 + \psi_1 = 1$.

Chapter 11
Probability Theory and Martingales

In the bulk of this book, we have avoided the rigorous formulation of stochastic processes used by probabilists. We have sacrificed the level of rigor in order to make the material accessible to applied mathematicians and biological physicists who tend not have a background in advanced probability theory. However, it is useful to have some exposure to the concepts and notation used by probabilists. Therefore, in this appendix, we give a very brief introduction to probability theory with an emphasis on martingales. There are a number of excellent textbooks on modern probability theory see for example [242, 681]. Both of these books tend to focus on discrete-time processes. However, the notation and concepts can be extended to continuous-time processes as detailed in [483].

11.1 Probability Spaces, Random Variables and Conditional Expectations

In appendix A we treated a continuous random variable X as a quantity that takes on a range of real values $x \in \mathbb{R}$, which are generated according to some probability distribution $F(x)$. Moreover, we assumed that there exists a probability density function p such that $dF(x) = p(x)dx$. We now turn to a more abstract definition of a random variable, which is used in more rigorous approaches to the theory of stochastic process. The starting point is a set of possible outcomes (of an experiment), which is denoted by the sample space Ω. An event is defined to be a subset A of Ω, which is some collection of single outcomes or elementary events $\omega \in \Omega$. In general not all subsets of Ω can be treated as events so that the set of events forms a subcollection $\mathscr{F}$ of all subsets. Within a probabilistic setting, this subcollection is required to be a so-called σ-algebra with the following properties:

1. $\emptyset \in \mathscr{F}$
2. If $A_1, A_2, \ldots \in \mathscr{F}$ then $\cup_{i=1}^{\infty} A_i \in \mathscr{F}$
3. If $A \in \mathscr{F}$ then $\Omega \backslash A \in \mathscr{F}$

P.C. Bressloff, *Stochastic Processes in Cell Biology*, Interdisciplinary Applied Mathematics 41, DOI 10.1007/978-3-319-08488-6_11

It can be shown that σ-algebras are closed under the operation of taking countable intersections. A *probability measure* $\mathbb{P}$ on $(\Omega, \mathscr{F})$ is a function $\mathbb{P} : \mathscr{F} \to [0,1]$ with

1. $\mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1$
2. If $A_i, A_j, \ldots \in \mathscr{F}$ with $A_i \cap A_j = \emptyset$, $i \neq j$, then

$$\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

The triple $(\Omega, \mathscr{F}, \mathbb{P})$ is called a *probability space*.

Given a function f on the sample space Ω, we can use the probability measure $\mathbb{P}$ to define the integral of this function over a set $A \in \mathscr{F}$ according to

$$f(A) = \int_A f(\omega) d\mathbb{P}(\omega).$$

If $f(\omega) = 1$ for all $\omega \in \Omega$, then $f(A) = \mathbb{P}(A)$. Note that for certain choices of σ-algebra, it is necessary to consider measures other than the standard Lebesgue measure. However, we will not consider this technicality here. A *random variable* is a function $X : \Omega \to \mathbb{R}$ such that

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathscr{F}, \quad \forall x \in \mathbb{R}.$$

If this condition holds, then X is said to be $\mathscr{F}$*-measurable*. If $X \in \mathbb{R}$, then we have a continuous random variable, whereas if X belongs to a countable set, then it is said to be a discrete random variable. The *distribution function* of a random variable X is the function $F : \mathbb{R} \to [0,1]$ given by

$$F(x) = \text{Prob}(X \leq x) = \mathbb{P}(X^{-1}(-\infty, x)),$$

where $X^{-1}(-\infty, x)$ is the set of events ω for which $X \leq x$. In the case of a continuous random variable, one can then express the expectation of random variable as an integral

$$\mathbb{E}(X) = \int_{\mathbb{R}} x dF(x),$$

where $dF(x) = \text{Prob}(x \leq X \leq x + dx)$. The probability density function (if exists) is then defined according to $dF(x) = \rho(x)dx$.

In the case of two random variables on $(\Omega, \mathscr{F}, \mathbb{P})$, we define the *conditional expectation* of Y given X by

$$\mathbb{E}(Y|X) = \int y \rho(y|X) dy,$$

where $\rho(y|X)$ is the conditional probability density with respect to X. The conditional expectation satisfies

$$\mathbb{E}(\mathbb{E}(Y|X)) \equiv \int \int y \rho(y|x) \rho(x) dy dx = \int y \rho_2(y,x) dy dx = \mathbb{E}(Y),$$

where ρ_2 is a joint probability density. Using a similar argument, one can also derive the tower property

$$\mathbb{E}(\mathbb{E}(Y|X_1,X_2)|X_1) = \mathbb{E}(Y|X_1).$$

An important generalization of the above is to construct the conditional expectation of a random variable Y with respect to a sub-σ-algebra $\mathscr{G} \subset \mathscr{F}$. Such a generalization plays an important role in the theory of martingales. First note that $\hat{Y} \equiv \mathbb{E}(Y|X)$ can be interpreted as a random variable satisfying

$$\int_{A(x)} \hat{Y}(\omega) d\mathbb{P}(\omega) = \int_{A(x)} Y(\omega) d\mathbb{P}(\omega) \quad \forall \quad x \in \mathbb{R},$$

where

$$A(x) = \{\omega \in \Omega : X(\omega) \leq x\} \subset \mathscr{F}.$$

The conditional expectation $\hat{Y} = \mathbb{E}(Y|\mathscr{G})$ for a general sub-σ-algebra $\mathscr{G}$ is then defined to be a random variable satisfying

$$\int_A \hat{Y}(\omega) d\mathbb{P}(\omega) = \int_A Y(\omega) d\mathbb{P}(\omega) \quad \forall \quad A \in \mathscr{G}.$$

It immediately follows from taking $A = \Omega$ that

$$\mathbb{E}(\mathbb{E}(Y|\mathscr{G})) = \mathbb{E}(Y).$$

Given two sub-σ-algebras $\mathscr{G}, \mathscr{H}$ with $\mathscr{G} \subset \mathscr{H} \subset \mathscr{F}$, one also has the tower property

$$\mathbb{E}(\mathbb{E}(Y|\mathscr{H})|\mathscr{G}) = \mathbb{E}(Y|\mathscr{G}).$$

11.2 Definitions and Examples of Martingales

Much of applied probability concerns establishing various limit theorems. Often this is achieved by showing that a particular sequence of random variables is a *martingale*. Such a sequence could be generated by a discrete-time stochastic process. (For the extension of martingales to continuous-time stochastic processes see [483].) In the following we fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and assume that there exists a given sequence $\mathscr{F}_n$, $n = 1, 2, \ldots$, of σ-algebras $\mathscr{F}_n \subseteq \mathscr{F}$.

Definition 11.1. A *filtration* on $(\Omega, \mathscr{F})$ is an increasing sequence

$$\mathscr{F}_1 \subseteq \mathscr{F}_1 \subseteq \mathscr{F}_2 \subseteq \ldots$$

of σ-algebras. A (discrete-time) stochastic process $(X_n)_{n\geq 1}$ is *adapted* to a filtration $(\mathscr{F}_n)_{n\geq 1}$ if and only if each X_n is $\mathscr{F}_n$-measurable, that is,

$$\{\omega \in \Omega : X_n(\omega) \leq x\} \in \mathscr{F}_n, \quad \forall x \in \mathbb{R}.$$

Example 11.1. The *canonical filtration* generated by a stochastic process (X_n) is given by

$$\mathscr{F}_n = \sigma(X_1, X_2, \ldots, X_n),$$

that is, the minimal event specifying the values of X_j, $j = 1, \ldots, n$. If the filtration is not specified explicitly, it will be assumed to be the canonical filtration. In general (X_n) is adapted to a filtration $(\mathscr{F}_n)$ if and only if $\sigma(X_1, X_2, \ldots, X_n) \subseteq \mathscr{F}_n$. Roughly speaking, as n increases, the statistical information about a larger class of random variables is included within the σ-algebra $\mathscr{F}_n$, as one might expect from the evolution of a discrete-time stochastic process.

Definition 11.2. A sequence of random variables $\{Y_n; n \geq 1\}$ on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is called a *martingale* with respect to the filtration $(\mathscr{F}_n)$ if and only if

(i) (Y_n) is adapted to $(\mathscr{F}_n)$

(ii) $\mathbb{E}(|Y_n|) < \infty$

(iii) $\mathbb{E}(Y_{n+1}|\mathscr{F}_n) = Y_n$

If $\mathscr{F}_n = \sigma(X_1, \ldots, X_n)$ for some other random sequence (X_n), then (Y_n) is said to be a martingale with respect to X. If (i) and (ii) hold but instead of (iii) we have

$$\mathbb{E}(Y_{n+1}|\mathscr{F}_n) \leq Y_n,$$

then (Y_n) is said to be a *supermartingale*. Similarly, if

$$\mathbb{E}(Y_{n+1}|\mathscr{F}_n) \geq Y_n,$$

then we have a *submartingale*.

Intuitively, the random sequence (Y_n) models the outcomes of a random process in time, whereas the filtration $(\mathscr{F}_n)$ specifies what is known at each step. Hence, if (Y_n) is a martingale, then we have at least partial information about the values $Y_1, \ldots, Y_n$, i.e., their probability distributions, and conditional on this information, the expected value of Y_{n+1} is equal (almost surely) to the observed value Y_n. Given the definition of a martingale, the following results hold:

(a) $\mathbb{E}(Y_n|\mathscr{F}_n) = Y_n$, which reflects the fact that Y_n is $\mathscr{F}_n$-measurable. It immediately follows that

$$\mathbb{E}(Y_{n+1} - Y_n|\mathscr{F}_n) = 0.$$

(b)

$$\mathbb{E}(Y_{n+k}|\mathscr{F}_n) = Y_n, \quad \forall \quad n \geq 1, k \geq 0,$$

which can be shown using induction. The case $k = 0$ holds from (a). Moreover the assertion for $k - 1$ implies that

$$\begin{aligned}\mathbb{E}(Y_{n+k}|\mathscr{F}_n) &= \mathbb{E}(\mathbb{E}(Y_{n+k}|\mathscr{F}_{n+k-1})|\mathscr{F}_n) \quad \text{from the tower property}\\ &= \mathbb{E}(Y_{n+k-1}|\mathscr{F}_n) \quad (Y_n) \text{ is a martingale}\\ &= Y_n.\end{aligned}$$

(c) Setting $n = 1$ in (b) shows that

$$\mathbb{E}(Y_n) = \mathbb{E}(Y_1) \quad \forall \quad n \geq 1,$$

that is, the expected outcome is independent of time.

Example 11.2. One simple example of a martingale is a 1D random walk, in which S_n is the position of the walker at the nth time step and $S_n = \sum_{j=1}^{n} X_j$ where X_j is the jth independent increment, with $X_j = 1$ (probability p) or $X_j = -1$ (probability $q = 1 - p$), and $S_0 = 0$. In this case, we take

$$\mathscr{F}_n = \sigma(X_1, \ldots, X_n) = \sigma(S_1, \ldots, S_n).$$

The result follows from

$$\begin{aligned}\mathbb{E}(S_{n+1}|X_1, X_2, \ldots, X_n) &= \mathbb{E}(S_n + X_{n+1}|X_1, X_2, \ldots, X_n)\\ &= \mathbb{E}(S_n|X_1, X_2, \ldots, X_n) + \mathbb{E}(X_{n+1}|X_1, X_2, \ldots, X_n)\\ &= S_n + p - q.\end{aligned}$$

Thus, $Y_n = S_n - n(p-q)$ defines a martingale with respect to X.

Example 11.3. Let $X_1, X_2, \ldots$ be a discrete-time Markov chain taking discrete values in some countable space Γ, with conditional probabilities

$$p_{ij} = \mathbb{P}(X_{n+1} = j|X_n = i).$$

Suppose that $\psi : \Gamma \to \mathbb{R}$ is a bounded function satisfying

$$\sum_{j \in \Gamma} p_{ij}\psi(j) = \psi(i), \quad \forall i \in \Gamma.$$

Then $Y_n = \psi(X_n)$ constitutes a martingale (with respect to X), since

$$\begin{aligned}\mathbb{E}(Y_{n+1}|X_1, \ldots X_n) &= \mathbb{E}(\psi(X_{n+1})|X_1, \ldots X_n)\\ &= \mathbb{E}(\psi(X_{n+1})|X_n) \quad \text{Markov property}\\ &= \sum_{j \in \Gamma} p_{X_n, j}\psi(j)\\ &= \psi(X_n) = Y_n.\end{aligned}$$

Example 11.4. Let X be a random variable on $(\Omega, \mathscr{F}, \mathbb{P})$. One can then define a sequence of conditional expectations with respect to a filtration $(\mathscr{F}_n)$ in $\mathscr{F}$ given by

$$X_n = \mathbb{E}(X|\mathscr{F}_n).$$

Note that X_n is itself a function of all the random variables contained within $\mathscr{F}_n$. From the tower property of conditional expectations and the fact that $\mathscr{F}_n \subseteq \mathscr{F}_{n+1}$, we have

$$\mathbb{E}(X_{n+1}|\mathscr{F}_n) = \mathbb{E}(\mathbb{E}(X|\mathscr{F}_{n+1})|\mathscr{F}_n) = \mathbb{E}(X|\mathscr{F}_n) = X_n.$$

11.3 Stopping Times

We have encountered many examples in cell biology where one needs to calculate a FPT. The latter is a particular example of a more abstract concept in probability theory known as a stopping time. This is based on the idea that the decision to stop a stochastic process at time m can be modeled as a random variable that is measurable with respect to $\mathscr{F}_m$, which represents the available information at time m.

Definition 11.3. A random variable $T : \Omega \to \{1,2,\ldots\} \cup \{\infty\}$ is called a stopping time with respect to the filtration $(\mathscr{F}_n)$ if and only if

$$\{T = n\} \in \mathscr{F}_n \quad \text{for any } n \geq 1.$$

Clearly $\{T \leq n\} \in \mathscr{F}_n$, since we have a filtration.

Example 11.5. A common example is the FPT for a discrete stochastic process (X_n) adapted to a filtration $(\mathscr{F}_n)$:

$$T_A = \min\{n \geq 0 : X_n \in A\}$$

for some measurable subset A of the state space. This follows from the observation that

$$\{T_A = m\} = \{X_1 \notin A\} \cap \ldots \cap \{X_{m-1} \notin A\} \cap \{X_m \in A\} \in \sigma(X_1, \ldots, X_n).$$

Note that $T_A = \infty$ if $X_n \notin A$ for all $n \geq 1$.

Let (X_n) be a sequence of random variables adapted to the filtration $(\mathscr{F}_n)$ in $\mathscr{F}$. Suppose that the sequence (X_n) is stopped according to a stopping time T with respect to $(\mathscr{F}_n)$. The sequence of random variables that will actually occur is (X_{T_n}) where

$$T_m = \min\{T, m\}.$$

In other words given a particular realization ω of the stochastic process

$$X_{T_m}(\omega) = \begin{cases} X_m(\omega) & \text{if } m \leq T(\omega) \\ X_T(\omega), & \text{if } m > T(\omega). \end{cases}$$

It can be shown that if (X_n) is a martingale with respect to the filtration $(\mathscr{F}_n)$, then so is the sequence (Y_n), where $Y_n = X_{T_n}$. If a martingale sequence (Y_n) is stopped at a fixed time n, the mean value satisfies $\mathbb{E}(Y_n) = \mathbb{E}(Y_1)$. An important issue within the

context of first passage properties of martingales is whether or not the expectation remains constant when the martingale is stopped after a random time T, that is, $\mathbb{E}(Y_T) = \mathbb{E}(Y_1)$. It turns out that such a result holds provided that T is a stopping time and there are a few additional constraints.

Theorem 11.1 (optional stopping theorem). *Let (X_n) be a martingale with respect to the filtration $(\mathscr{F}_n)$ and let T be a stopping time. Then $\mathbb{E}(X_T) = \mathbb{E}(X_1)$ if:*

(a) $\mathbb{P}(T < \infty) = 1$

(b) $\mathbb{E}(X_T) < \infty$

(c) $\mathbb{E}(X_n I_{\{T>n\}}) \to 0$ *as* $n \to \infty$

Here I_A for $A \in \mathscr{F}$ is an indicator function with $I_A(\omega) = 1$ if $\omega \in A$ and $I_A(\omega) = 0$ if $\omega \notin A$.

Proof. First note that the difference between X_T and X_{T_n} is zero if $T < n$. Therefore, we have the decomposition

$$X_T - X_{T_n} = (X_T - X_n) I_{\{T>n\}}.$$

Taking the expectation of both sides,

$$\mathbb{E}(X_T) - \mathbb{E}(X_{T_n}) = \mathbb{E}(X_T I_{\{T>n\}}) - \mathbb{E}(X_n I_{\{T>n\}}).$$

Since (X_{T_n}) is a martingale, it follows that $\mathbb{E}(X_{T_n}) = \mathbb{E}(X_1)$. Moreover, in the limit $n \to \infty$ the last term on the right-hand side vanishes due to condition (c). Also,

$$\mathbb{E}(X_T I_{\{T>n\}}) = \sum_{k=n+1}^{\infty} \mathbb{E}(X_T I_{\{T=k\}}),$$

which is the tail of the convergent series

$$\mathbb{E}(X_T) = \sum_{k=1}^{\infty} \mathbb{E}(X_T I_{\{T=k\}}) < \infty.$$

Thus, taking the limit $n \to \infty$ establishes that $\mathbb{E}(X_T) = \mathbb{E}(X_1)$.

Example 11.6. Let us return to the random walk of Example 11.2 for $p = q = 1/2$, in which case (S_n) is a martingale. Let a, b be positive integers. Introduce the stopping time

$$T = \min\{n : S_n = -a \text{ or } S_n = b\}$$

which is the first time that the walker visits the endpoints $-a, b$ of the interval $[-a, b]$ with $a < 0 < b$. It can be shown that T satisfies the conditions of the optional stopping theorem. Let p_a be the splitting probability that the walker reaches $-a$ before it reaches b. By the optional stopping theorem,

$$\mathbb{E}(S_T) = -a p_a + b(1 - p_a) = \mathbb{E}(S_0) = 0.$$

This then determines p_a:

$$p_a = \frac{b}{a+b}.$$

Using the fact that $Y_n = S_n^2 - n$ also generates a martingale sequence, the optional stopping theorem implies

$$0 = \mathbb{E}(S_T^2 - T) = a^2 p_a + b^2(1-p_a) - \mathbb{E}(T),$$

that is,

$$\mathbb{E}(T) = a^2 p_a + b^2(1-p_a) = \frac{a^2 b}{a+b} + \frac{ab^2}{a+b} = ab.$$

11.4 Martingale Convergence Theorems

Let $X_1, X_2, \ldots, X$ be random variables on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$. There are four basic ways of interpreting the meaning of $X_n \to X$.

1. $X_n \xrightarrow{\text{a.s.}} X$ almost surely if

$$\{\omega \in \Omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\}$$

is an event of probability one.

2. $X_n \xrightarrow{\text{r}} X$ in rth mean, $r \geq 1$, if

$$\mathbb{E}(|X_n - X|^r) \to 0 \text{ as } n \to \infty.$$

3. $X_n \xrightarrow{\text{p}} X$ in probability if

$$\mathbb{P}(|X_n - X| > \varepsilon) \to 0 \text{ as } n \to \infty \quad \forall \varepsilon > 0.$$

4. $X_n \xrightarrow{\text{D}} X$ in distribution if

$$\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x) \text{ as } n \to \infty$$

for all $x \in \mathbb{R}$ satisfying $\mathbb{P}(X = x) = 0$, where $\mathbb{P}(X \leq x)$ is continuous.

Given these definitions, the following implications hold:

$$X_n \xrightarrow{\text{a.s.}} X \text{ or } X_n \xrightarrow{\text{r}} X \implies X_n \xrightarrow{\text{p}} X \implies X_n \xrightarrow{\text{D}} X.$$

One of the useful properties of martingales is that one can prove various limit theorems depending on the particular choice of convergence. Here we consider mean-square or L^2 convergence, which is the easiest to establish using standard analysis. Let (X_n) be a martingale with respect to the filtration $(\mathscr{F}_n)$ and assume

square integrability $\mathbb{E}(X_n^2) < \infty$ for all $n \geq 1$ and set $X_0 = 0$. One useful property of square-integrable martingales is that the increments $Y_n = X_n - X_{n-1}$ have zero mean and are uncorrelated in L^2. That is, since (X_n) is a martingale

$$\mathbb{E}(Y_n|\mathscr{F}_{n-1}) = \mathbb{E}(X_n|\mathscr{F}_{n-1}) - \mathbb{E}(X_{n-1}|\mathscr{F}_{n-1}) = X_{n-1} - X_{n-1} = 0$$

for all $n \geq 1$. Hence, from the properties of conditional expectations, $\mathbb{E}(Y_n) = \mathbb{E}(\mathbb{E}(Y_n|\mathscr{F}_{n-1})) = 0$, and

$$\mathbb{E}(Y_m Y_n) = \mathbb{E}(Y_m \mathbb{E}(Y_n|\mathscr{F}_{n-1})) = 0, \quad \text{for } m < n.$$

Theorem 11.2 (L^2 martingale convergence theorem). *If (X_n) is a nonnegative martingale, such that $\mathbb{E}(X_n) < \infty$ for all n, then there exists a random variable X with $\mathbb{E}(X) < \infty$ and satisfying the following limits:*

1. $\lim_{n\to\infty} X_n = X$ *almost surely.*
2. *If $\mathbb{E}(X_n^2) < M < \infty$ for some M and all n, then X_n also converges to X in the L^2 sense. In particular, $var(X) = \lim_{n\to\infty} var(X_n)$.*

Proof (part 2). First note that for $m \leq n$

$$\begin{aligned}\mathbb{E}(X_n^2) - \mathbb{E}(X_m^2) &= \mathbb{E}((X_n - X_m)(X_n + X_m)) = \mathbb{E}((X_n - X_m)^2) + 2\mathbb{E}(X_m(X_n - X_m)) \\ &= \mathbb{E}((X_n - X_m)^2).\end{aligned}$$

It follows that $\mathbb{E}(X_n^2)$ is an increasing function of n. Since this sequence is bounded, it is a Cauchy sequence and thus (X_n) is a Cauchy sequence in L^2. Completeness of L^2 establishes that (M_n) converges.

11.5 The Galton–Watson Branching Process

We now illustrate an application of the martingale convergence theorem by considering the Galton–Watson branching process, which is often used to model cell proliferation (see Chap. 3 of Kimmel and Axelrod [334]). The process starts with a single ancestor who produces a random number of progeny according to a given probability distribution. Each member of the first generation behaves independently and produces second-generation progeny in an identical fashion to the ancestor. Iterating this procedure leads to the Galton–Walton branching process. Let Z_n denote the number of members (events) of the nth generation (time step). Each member of the nth generation gives birth to a family, which could be empty, of members of the $(n+1)$th generation with the following assumption: the family sizes of the individuals of the branching process are independent identically distributed random variables. An example realization of a branching process is shown in Fig. 11.1.

11.5.1 Generating Functions and Basic Properties

A useful method for analyzing a branching process is to use generating functions. Let $G_n(s) = \mathbb{E}(s^{Z_n})$ be the generating function of the random variable Z_n with probability $P_n(m) = \text{Prob}(Z_n = m)$. Each member of the $(n+1)$th generation has a unique ancestor in the nth generation such that

$$Z_{n+1} = X_1 + X_2 + \ldots + X_{Z_n},$$

where X_i is the size of the family produced by the ith member of the mth generation. It follows that

$$\begin{aligned} G_{n+1}(s) &= \mathbb{E}(s^{Z_{n+1}}) = \mathbb{E}(s^{X_1+\ldots+X_{Z_n}}) \\ &= \sum_{n=0}^{\infty} \mathbb{E}(s^{X_1+\ldots+X_m} | Z_n = m) P_n(m) \\ &= \sum_{n=0}^{\infty} \mathbb{E}(s^{X_1} s^{X_2} \ldots^{X_m} | Z_n = m) P_n(m) \\ &= \sum_{n=0}^{\infty} \left[\prod_{j=1}^{m} \mathbb{E}(s^{X_j}) \right] P_n(m) \\ &= \sum_{n=0}^{\infty} [G_1(s)]^m P_n(m) = G_n(G_1(s)). \end{aligned}$$

Iterating this resulting and dropping the subscript on G_1, we have the recurrence relation

$$G_n(s) = G_{n-1}(G(s)) = G(G(\ldots(G(s))\ldots)). \tag{11.5.1}$$

Let $\mu = \mathbb{E}(Z_1)$ and $\sigma^2 = \text{var}(Z_1)$. In order to determine the mean and variance of Z_n, we use the recursive structure of the generating functions. First,

$$\mathbb{E}(Z_n) = G_n'(1) = \frac{d}{ds} G(G_{n-1}(s)) \bigg|_{s=1} = G'(1) G_{n-1}'(s) = \mu G_{n-1}'(s).$$

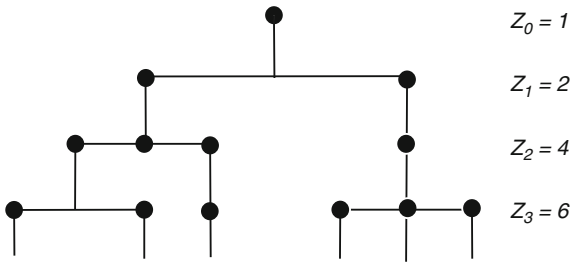


Fig. 11.1: Illustration of a branching process with three generations

Iterating this result shows that

$$\mathbb{E}(Z_n) = \mu^n. \tag{11.5.2}$$

Similarly,

$$\mathbb{E}(Z_n(Z_n-1)) = G_n''(1) = G''(1)G_{n-1}'(1)^2 + G'(1)G_{n-1}''(1).$$

This gives the iterative result

$$\operatorname{var}(Z_n) = \sigma^2\mu^{2n-2} + \mu\operatorname{var}(Z_{n-1}),$$

from which one finds that

$$\operatorname{var}(Z_n) = \begin{cases} n\sigma^2 & \text{if } \mu = 1 \\ \dfrac{\sigma^2(\mu^n-1)\mu^{n-1}}{\mu-1} & \text{if } \mu \neq 1. \end{cases} \tag{11.5.3}$$

Let T_n be the total number of individuals up to and including the nth generation. Then

$$\begin{aligned} \mathbb{E}(T_n) &= \mathbb{E}(Z_0 + Z_1 + Z_2 + \ldots + Z_n) \\ &= 1 + \mathbb{E}(Z_1) + \mathbb{E}(Z_2) + \ldots + \mathbb{E}(Z_n) \\ &= 1 + \mu + \mu^2 + \ldots + \mu^n \\ &= \begin{cases} \frac{\mu^{n+1}-1}{\mu-1}, & \mu \neq 1, \\ n+1, & \mu = 1. \end{cases} \end{aligned}$$

It follows that

$$\lim_{n\to\infty} \mathbb{E}(T_n) = \begin{cases} \infty, & \mu \geq 1, \\ \frac{1}{1-\mu}, & \mu < 1. \end{cases}$$

Let $H_n(s) = \mathbb{E}(s^{T_n})$ be the generating function for the random variable T_n. The generating functions satisfy the recurrence relation

$$H_{n+1}(s) = sG(H_n(s)). \tag{11.5.4}$$

11.5.2 Extinction and Criticality

An important property of a branching process is whether or not it eventually becomes extinct, that is, $Z_n = 0$ for some finite n. This motivates the classification of a branching process is in terms of the asymptotic properties of the mean number of progeny $\mathbb{E}(Z_n)$. Since $\mathbb{E}(Z_n) = \mu^n$ for a Galton–Watson process, we see that it grows geometrically if $\mu > 1$, stays constant if $\mu = 1$, and decays geometrically if $\mu < 1$. These three cases are labeled supercritical, critical, and subcritical, respectively:

$$\begin{aligned} \mu > 1 \text{ (supercritical)} &\implies \lim_{n\to\infty} \mathbb{E}(Z_n) = \infty \\ \mu = 1 \text{ (critical)} &\implies \mathbb{E}(Z_n) = 1 \\ \mu < 1 \text{ (subcritical)} &\implies \lim_{n\to\infty} \mathbb{E}(Z_n) = 0. \end{aligned}$$

The relationship between criticality and the probability of extinction η, where

$$\eta = \lim_{n\to\infty} \text{Prob}(Z_n = 0), \tag{11.5.5}$$

is quite subtle. First note that

$$\eta_n \equiv \text{Prob}(Z_n = 0) = G_n(0) = G(G_{n-1}(0)) = G(\eta_{n-1}).$$

Taking the limit $n \to \infty$ shows that η is a root of the equation $\eta = G(\eta)$. Moreover, if ψ is any nonnegative root of $s = G(s)$, then $\eta \le \psi$. This follows from the fact that G is a non-decreasing function on $[0, 1]$ so

$$\eta_1 = G(0) \le G(\psi) = \psi, \eta_2 = G(\eta_1) \le G(\psi) = \psi, \ldots,$$

so by induction $\eta_n \le \psi$ for all $n \ge 0$. Hence η is the smallest nonnegative root of the equation $s = G(s)$. The value of η can now be determined graphically by noting that $G(1) = 1$ and $G(s)$ is a convex function of s (see Fig. 11.2). The latter is a consequence of the result

$$G''(s) = \mathbb{E}(Z_1(Z_1 - 1)s^{Z_1 - 2}) \ge 0, \quad \forall \quad s \ge 0.$$

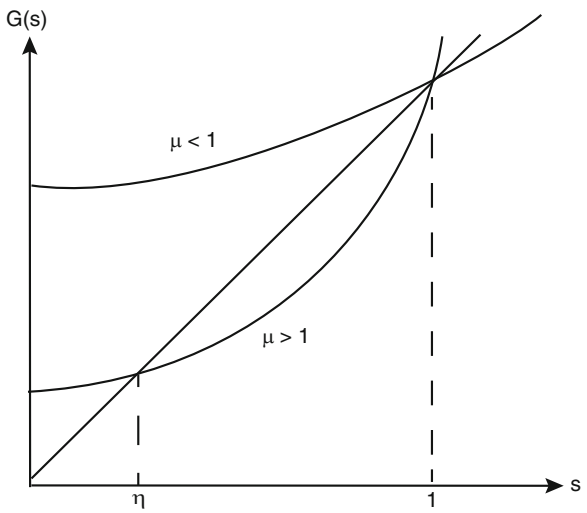


Fig. 11.2: Graphical construction of the probability of extinction η. The convex function $G(s)$, $s \ge 0$, intersects the diagonal twice for $\mu > 1$ but only once when $\mu \le 1$

Hence, if $\mu = G'(1) > 1$, then there exists a root $\eta = G(\eta)$ with $0 \leq \eta < 1$, whereas if $\mu = G'(1) \leq 1$, then $\eta = 1$ (the two roots coincide). We see that the critical process is counterintuitive, since the process becomes extinct almost surely although the mean $\mathbb{E}(Z_n) = 1$ for all $n \geq 0$.

Example 11.7. For the sake of illustration, consider a geometric branching process where the distribution of family sizes is given by $\text{Prob}(Z_1 = k) \equiv f(k) = qp^k$ with $q = 1 - p$. In this case, one can calculate the generating function and other quantities explicitly:

$$G(s) = q(1 - ps)^{-1}, \quad \mu = \frac{p}{q}, \quad \sigma^2 = \frac{p^2}{q^2} + \frac{p}{q}. \tag{11.5.6}$$

Moreover, it can be shown by induction that

$$G_n(s) = \begin{cases} \dfrac{n - (n-1)s}{n + 1 - ns} & \text{if } p = q = \frac{1}{2}, \\ \dfrac{q[p^n - q^n - ps(p^{n-1} - q^{n-1})]}{p^{n+1} - q^{n+1} - ps(p^n - q^n)} & \text{if } p \neq q. \end{cases} \tag{11.5.7}$$

It follows that

$$\text{Prob}(Z_n = 0) = G_n(0) = \begin{cases} \dfrac{n}{n+1} & \text{if } p = q, \\ \dfrac{q(p^n - q^n)}{p^{n+1} - q^{n+1}} & \text{if } p \neq q, \end{cases} \tag{11.5.8}$$

and, hence, $\eta = 1$ if $p \leq q$ and $\eta = q/p$ if $p > q$. We conclude that for a geometric branching process, extinction occurs almost surely if $\mathbb{E}(Z_1) = \mu = p/q \leq 1$; otherwise there is a finite probability of persistent growth. We can identify the regime $p < q$ as subcritical, the regime $p > q$ as supercritical, and the point $p = q = 1/2$ as critical.

11.5.3 Asymptotic Properties

The asymptotic properties of the Galton–Watson process can be analyzed using the convergence properties of martingales, as summarized in Theorem 11.2. The applicability of this theorem to the Galton–Watson process follows from the fact that $W_n = Z_n/\mu^n$ is a martingale. That is, since the branching process is a Markov chain,

$$\mathbb{E}(Z_{n+1}|Z_n, \ldots, Z_0) = \mathbb{E}(Z_{n+1}|Z_n),$$

and thus

$$\mathbb{E}(W_{n+1}|W_n, \ldots, W_0) = \mathbb{E}(W_{n+1}|W_n).$$

Moreover, since Z_{n+1} is the sum of Z_n independent families, we have

$$\mathbb{E}(Z_{n+1}|Z_n) = \mu Z_n.$$

Similarly,

$$\mathbb{E}(W_{n+1}|W_n) = \mu^{-(n+1)}\mathbb{E}(Z_{n+1}|Z_n) = \frac{Z_n}{\mu^n} = W_n.$$

We deduce that (W_n) is a martingale:

$$\mathbb{E}(W_{n+1}|W_n, \ldots, W_0) = W_n.$$

The convergence theorem thus establishes that there exists a random variable W for which

$$\lim_{n\to\infty} W_n = W \quad \text{almost surely.}$$

Since $\eta = 1$ in the critical and subcritical cases, it follows that $W \equiv 0$. Hence, we assume the process is supercritical. Now introduce the discrete Laplace transform of W_n given by $\phi_n(s) = \mathbb{E}(\mathrm{e}^{-sW_n})$. We have

$$\begin{aligned}\phi_n(s) &= \mathbb{E}(\mathrm{e}^{-sW_n}) = \mathbb{E}[(\mathrm{e}^{-s/\mu^n})^{Z_n}] \\ &= G_n(\mathrm{e}^{-s/\mu^n}) = G(G_{n-1}(\mathrm{e}^{-s/\mu^n})) \\ &= G(\phi_{n-1}(s/\mu)).\end{aligned}$$

Since $W_n \to W$ almost surely implies $W_n \to W$ in distribution, taking the limit $n \to \infty$ shows that $\phi_n(s) \to \phi(s)$, with ϕ satisfying the so-called Abel's equation

$$\phi(s) = G\left[\phi(s/\mu)\right]. \tag{11.5.9}$$

There are also nontrivial limit theorems for subcritical and critical process conditions on nonextinction. We simply quote the theorems here (see Athreya and Ney [16] for proofs).

Theorem 11.3. *If $\mu < 1$ (subcritical), then* $\mathrm{Prob}(Z_n|Z_n > 0)$ *converges as $n \to \infty$ to a probability distribution whose generating function $B(s)$ satisfies*

$$B[G(s)] = \mu B(s) + 1 - \mu,$$

and the probability of nonextinction has the asymptotic form

$$1 - \eta_n \sim \frac{\mu^n}{B'(1)}, \quad n \to \infty.$$

Theorem 11.4. *If $\mu = 1$ (critical) and $\sigma^2 = var(Z_1) < \infty$, then*

$$\lim_{n\to\infty} \mathrm{Prob}(Z_n/n > z|Z_n > 0) = \mathrm{e}^{-2z/\sigma^2}, \quad z \geq 0.$$

11.5.4 Application to Gene Amplification and Drug Resistance

There are many applications of the Galton–Watson process to models of cell proliferation and population genetics, as reviewed in [334]. Here we illustrate the theory by considering the particular example of gene amplification [333, 334]. This refers to the increase in the number of copies of a gene through successive cell generations. One important example is the amplification of genes coding for the enzyme DHFR, which has been associated with cellular resistance to the anticancer drug methotrexate (MTX). A Galton–Watson process can be used to model the number Z_n of DHFR genes in a cell randomly selected from cell progeny in the nth generation of repeated cycles of cell replication and division (see Fig. 11.3). During the lifetime of a cell, each DHFR gene is either replicated with probability a or not replicated with probability $1-a$. In the former case, at the time of cell division, DHFR is assigned to one of the two daughter cells with probability $1/2$. On the other hand, if replication occurs, then during cell division both copies are assigned to one daughter cell with probability $\alpha/2$ or each daughter cell receives one copy with probability $1-\alpha$. It follows that for a randomly selected cell of the first generation, given $Z_0 = 1$, we have

$$\text{Prob}(Z_1 = 0) = (1-a)/2 + a\alpha/2 \equiv A_0, \quad \text{Prob}(Z_1 = 1) = 1 - A_1 - A_2,$$
$$\text{Prob}(Z_1 = 2) = a\alpha/2 = A_2.$$

Hence

$$G(s) = \mathbb{E}(s^{Z_1}) = A_0 + (1 - A_0 - A_2)s + A_2 s^2.$$

It is assumed that in the absence of selection, DHFR gradually disappears from the cell population so that the branching process is subcritical. This means that

$$\mu \equiv G'(1) = 1 - A_0 + A_2 < 1 \implies A_2 < A_0.$$

A cell is said to be resistant if it contains at least one copy of the DHFR gene; otherwise it is called sensitive. Suppose that cells are initially cultured in a medium rich in MTX so that they develop drug resistance in the sense that all cells have at least one copy of DHFR. After N generations the distribution of copy numbers is given by $\text{Prob}(Z_N | Z_N > 0)$. Moreover, since the probability of nonextinction scales as μ^n for large n, it follows that the probability of a daughter cell to be resistant is μ so that the number of resistant cells grows on average by 2μ per generation (assuming $1/2 < \mu < 1$). Now suppose that at the Nth generation the drug-resistant cells are placed in a drug-free medium so sensitive cells also proliferate. Let $R(n)$ and $S(n)$ denote the number of resistant and sensitive cells after n cycles within the drug-free medium with $S(0) = 0$. Using the fact that $R(n) = (2\mu)^n R(0)$ and $S(n) + R(n) = 2^n R(0)$, we see that the fraction of drug-resistant cells scales as

$$\frac{r(n)}{r(0)} = \mu^n \to 0 \text{ as } n \to \infty, \quad r(n) = \frac{R(n)}{R(n) + S(n)}.$$

Meanwhile the distribution of copy numbers among the drug-resistant population is preserved after normalization, consistent with experimental observations.

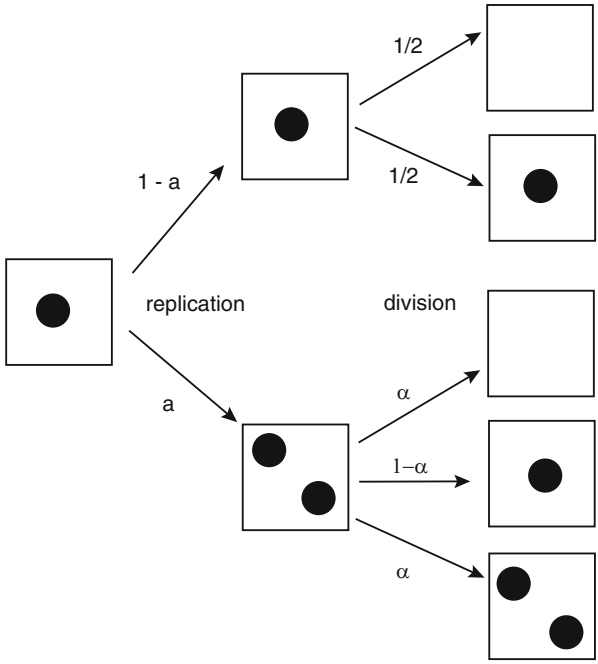


Fig. 11.3: Schematic illustration of a branching process model of gene amplification. See text for details

11.6 Modeling Chemical Reaction Networks as Counting Processes

As our final example illustrating the application of martingales, we turn to a stochastic formulation of biochemical reaction networks based on counting processes, which provides a framework for applying rigorous probabilistic methods such as the theory of continuous-time martingales [9, 361]. The reaction network is treated as a continuous-time Markov chain with reactions corresponding to transitions along the chain. The number of occurrences of each reaction is modeled as a counting process, which is itself represented in terms of a scaled Poisson process. This so-called random time-change representation yields a stochastic equation for the Markov chain. Our presentation follows closely the introductory review by Anderson and Kurtz [9].

11.6.1 Poisson Processes and Counting Processes

Let us first briefly recall properties of a Poisson process $Y(t)$, which represents the number of events or observations up to time t with the following properties:

1. Observations occur one at a time.
2. The number of observations in disjoint time intervals are independent random variables. That is, if $t_0 < t_1 < \ldots < t_m$, then $Y(t_k) - Y(t_{k-1})$, $k = 1, \ldots, m$ are independent random variables.
3. The distribution of $Y(t+a) - Y(t)$ is independent of t.

Under the above assumptions it can be shown that there exists a constant $\lambda > 0$ such that, for $t < s$, $Y(s) - Y(t)$ is Poisson distributed with rate λ:

$$\mathbb{P}[Y(s) - Y(t) = k] = \frac{\lambda(s-t)^k}{k!} \mathrm{e}^{-\lambda(s-t)}.$$

In the following we will take $Y(t)$ to be a unit rate Poisson process ($\lambda = 1$) and denote a Poisson process with rate λ by the time-change representation $Y(\lambda t)$. Suppose that $\mathscr{F}_t$ represents the information obtained by observing $Y(\lambda s)$ for $s \leq t$. It follows that

$$\mathbb{P}[Y(\lambda(t+\Delta t)) - Y(\lambda t) = 1 | \mathscr{F}_t] = \mathbb{P}[Y(\lambda(t+\Delta t)) - Y(\lambda t) = 1] = \lambda \Delta t \mathrm{e}^{-\lambda \Delta t}.$$

Taking the limit $\Delta t \to 0$, we have the formal limit

$$\mathbb{P}[dY(\lambda t) = 1 | \mathscr{F}_t] \equiv \mathbb{E}[dY(\lambda t) | \mathscr{F}_t] = \lambda dt. \tag{11.6.1}$$

Thus one can interpret the rate λ as the expected number of jumps per unit time (transition intensity or propensity). The usefulness of this definition of λ is that it is intuitive and can easily be generalized to a large class of counting processes,[1] including those that keep track of the number of single-step chemical reactions (see below). Moreover, it forms the starting point of a mathematically rigorous theory of counting process that involves the theory of continuous-time martingales. Indeed, a more precise version of the formal relation (11.6.1) is that in terms of the filtration $\{\mathscr{F}_t, t \geq 0\}$, the stochastic process

$$M(t) = Y(\lambda t) - \lambda t$$

is a martingale. This follows from

$$\mathbb{E}[Y(\lambda[t+s]) - Y(\lambda t) | \mathscr{F}_t)] = \lambda s \implies \mathbb{E}[Y(\lambda[t+s]) | \mathscr{F}_t)] = Y(\lambda t) - \lambda t + \lambda(t-s),$$

[1] A counting process is a stochastic process $\{N(t), t \geq 0\}$ satisfying the following properties: $N(t) \geq 0$; $N(t)$ is an integer; if $s \leq t$, then $N(s) \leq N(t)$ (a positive, increasing integer).

that is,

$$\mathbb{E}[M(t+s)|\mathscr{F}_t] = M(t)$$

for all $s,t \geq 0$.

The above results generalize to a Poisson process with a time-dependent rate function $\lambda(t)$ so that the counting process $N(t) = Y(\lambda(t))$:

1. The random variable $Y(\lambda(t)) - Y(\lambda(s))$ is independent of $\mathscr{F}_s$ for $t > s$.
2. The conditional distribution of the increments is given by

$$\mathbb{P}[Y(\lambda(t)) - Y(\lambda(s)) = k|\mathscr{F}_s] = \frac{\Lambda_{s,t}^k}{k!}\mathrm{e}^{-\Lambda_{s,t}},$$

   where

$$\Lambda_{s,t} = \int_s^t \lambda(u)du.$$

3. The stochastic process

$$M(t) = Y(\lambda(t)) - \int_0^t \lambda(u)du$$

   is a martingale with respect to the filtration $\{\mathscr{F}_t\}$.

The converse relation also holds, namely, if $M(t)$ is a martingale with respect to $\mathscr{F}_t$, then $N(t) = Y(\lambda(t))$, that is, $N(t)$ is a Poisson process (Watanabe theorem).

In order to develop a mathematical theory of chemical reaction networks, it is first necessary to consider a more general counting process $N(t)$ in which the intensity $\lambda(t;N)$ is itself a stochastic process adapted to the filtration $\{\mathscr{F}_t\}$. In other words, if $\mathscr{F}_t$ represents all information about the counting process up to time t, then $\lambda(t;N)$ is specified, that is, it is *non-anticipating*. For concreteness, take $\lambda(t;N) = \lambda(N(t))$ and consider the counting process given by the solution to the stochastic equation

$$N(t) = Y\left(\int_0^t \lambda(N(s))ds\right), \tag{11.6.2}$$

with $\int_0^t \lambda(N(s))ds < \infty$ for all $t \geq 0$. Note that in the infinitesimal interval $(t,t+\Delta t]$,

$$\begin{aligned}\mathbb{P}[N(t+\Delta t) > N(t)|\mathscr{F}_t] &= 1 - \mathbb{P}[N(t+\Delta t) = N(t)|\mathscr{F}_t] \\ &= 1 - \mathbb{P}\left[Y\left(\int_0^{t+\Delta t} \lambda(N(s))ds\right) - Y\left(\int_0^t \lambda(N(s))ds\right) = 0\right] \\ &= 1 - \mathrm{e}^{-\lambda(N(t))\Delta t} \approx \lambda(N(t))\Delta t.\end{aligned}$$

Thus $\lambda(N(t))$ can still be identified as a transition intensity. The relationship to martingales is now a little more involved. First, define the stochastic jump times

$$\tau_k = \inf\{t, t \geq 0 | N(t) \geq k\}$$

with τ_k the time of the kth jump, and introduce the notation

$$\tau_k \wedge t = \min(\tau_k, t).$$

Define the following set of stochastic processes:

$$M_k(t) = N(t \wedge \tau_k) - \int_0^{t\wedge\tau_k} \lambda(N(s))ds. \tag{11.6.3}$$

Given the counting process $N(t)$ satisfying equation (11.6.2), $M_k(t)$ is a martingale for all $k \geq 0$, that is,

$$\mathbb{E}[M_k(t+s)|\mathscr{F}_t] = M_k(t), \quad t,s \geq 0.$$

Moreover, if $\mathbb{E}[N(t)] < \infty$ for all $t \geq 0$, then $\lim_{k\to\infty} \tau_k \equiv \tau_\infty = \infty$ and

$$M(t) \equiv \lim_{k\to\infty} M_k(t) = N(t) - \int_0^t \lambda(N(u))du$$

is a martingale. Since $M(0) = 0$ (assuming $N(0) = 0$), it follows that $\mathbb{E}[M(t)] = 0$ for all $t \geq 0$, that is,

$$\mathbb{E}\left[Y\left(\int_0^t \lambda(N(s))ds\right)\right] = \mathbb{E}\left[\int_0^t \lambda(N(u))du\right].$$

Again, the converse relation holds: suppose that $N(t)$ is a counting process with transition intensity $\lambda(N(t))$ such that for each k, $M_k(t)$ is a martingale. Then $N(t)$ is the solution to the stochastic equation (11.6.2).

Martingales of a counting process.

We sketch a proof that $M_k(t)$ is a martingale for the counting process $N(t)$ satisfying the stochastic equation (11.6.2). Suppose that $\tau_{k-1} < t < \tau_k$. Since τ_{k-1} is adapted to $\mathscr{F}_t$, it follows that τ_{k-1} is known when conditioning expectations with respect to $\mathscr{F}_t$. On the other hand, τ_k is a random variable. It is clear from the definition of $M_k(t)$ that for all $j \leq k-1$

$$\mathbb{E}[M_j(t+s)|\mathscr{F}_t] = N(\tau_j) - \int_0^{\tau_j} \lambda(N(u))du = M_j(t).$$

In the case $j = k$, we have to consider two cases: (i) $\tau_k < t+s$, which means that at least one jump occurs in the interval $(t, t+s]$ and $N(t+s) - N(t) > 0$; (ii) $\tau_k \geq t+s$, which means that no jump occurs in $(t, t+s)$ and $N(t+s) = N(t)$. Therefore

$$\begin{aligned}
\mathbb{E}[M_k(t+s)|\mathscr{F}_t] &= \int_{t+s}^{\infty} \mathbb{P}(\tau_k|\mathscr{F}_t)\left[N(t+s) - \int_0^{t+s} \lambda(N(u))du\right] d\tau_k \\
&\quad + \int_t^{t+s} \mathbb{P}(\tau_k|\mathscr{F}_t)\left[N(\tau_k) - \int_0^{\tau_k} \lambda(N(u))du\right] d\tau_k \\
&= \int_{t+s}^{\infty} \mathbb{P}(\tau_k|\mathscr{F}_t)\left[N(t) - \int_0^{t} \lambda(N(u))du - \int_t^{t+s} \lambda(N(u))du\right] d\tau_k \\
&+ \int_t^{t+s} \mathbb{P}(\tau_k|\mathscr{F}_t)\left[N(t) + 1 - \int_0^{t} \lambda(N(u))du - \int_t^{\tau_k} \lambda(N(u))du\right] d\tau_k \\
&= \int_{t+s}^{\infty} \mathbb{P}(\tau_k|\mathscr{F}_t)\left[M_k(t) - \int_t^{t+s} \lambda(N(u))du\right] d\tau_k \\
&+ \int_t^{s+t} \mathbb{P}(\tau_k|\mathscr{F}_t)\left[M_k(t) + 1 - \int_t^{\tau_k} \lambda(N(u))du\right] d\tau_k.
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbb{E}[M_k(t+s)|\mathscr{F}_t] - M_k(t) &= -\int_{s+t}^{\infty} \mathbb{P}(\tau_k|\mathscr{F}_t) d\tau_k \left[\int_t^{t+s} \lambda(N(u))du\right] \\
&\quad + \int_t^{s+t} \mathbb{P}(\tau_k|\mathscr{F}_t)\left[1 - \int_t^{\tau_k} \lambda(N(u))du\right] d\tau_k,
\end{aligned}$$

where

$$\mathbb{P}(\tau_k|\mathscr{F}_t) = C\exp\left(-\int_t^{\tau_k} \lambda(N(u))du\right)$$

and C is a normalization factor. Conditioning in $\mathscr{F}_t$ and the absence of any subsequent jumps means that $\lambda(N(u))$ can be treated as a constant λ_0. Then

$$\begin{aligned}
\mathbb{E}[M_k(t+s)|\mathscr{F}_t] - M_k(t) &= -Cs\lambda_0 \int_{s+t}^{\infty} \mathrm{e}^{-\lambda_0(\tau_k - t)} d\tau_k \\
&\quad + C\int_t^{s+t} \mathrm{e}^{-\lambda_0(\tau_k - t)}\left[1 - (\tau_k - t)\lambda_0\right] d\tau_k \\
&= -Cs\mathrm{e}^{-\lambda_0 s} + C\frac{1 - \mathrm{e}^{-\lambda_0 s}}{\lambda_0} + C\lambda_0 \frac{d}{d\lambda_0}\frac{1 - \mathrm{e}^{-\lambda_0 s}}{\lambda_0} \\
&= 0.
\end{aligned}$$

Thus

$$\mathbb{E}[M_k(t+s)|\mathscr{F}_t] = M_k(t) \text{ for all } 0 \le t < \tau_k, \quad s \ge 0.$$

Since $\mathbb{E}[M_k(t+s)|\mathscr{F}_t] = M_k(t)$ for all $t > \tau_k$, it follows that $M_k(t)$ is a martingale. This result holds for all $k \ge 0$.

The above martingale formulation can be used to derive the master equation for the counting process. Introduce an arbitrary, bounded function f on $\mathbb{Z}$. Suppose that $N(t) = n$ and τ_k is the nth jump time. Then

$$\begin{aligned} f(N(t)) &= f(0) + \sum_{k=1}^{n} [f(k) - f(k-1)] = f(0) + \sum_{k=1}^{n} [f(N(\tau_k)) - f(N(\tau_{k-1}))] \\ &= f(0) + \sum_{k=1}^{n} [f(N(\tau_{k-1}) + 1) - f(N(\tau_{k-1}))] \\ &= f(0) + \sum_{k=1}^{\infty} \int_0^t [f(N(s) + 1) - f(N(s))] \delta(s - \tau_{k-1}) ds, \\ &= f(0) + \sum_{k=1}^{\infty} \int_0^t [f(N(s-) + 1) - f(N(s-))] \delta(s - \tau_k) ds, \end{aligned}$$

which is independent of the particular value n. Formally speaking,

$$\sum_{k=1}^{\infty} \delta(s - \tau_k) = dN(t),$$

so that

$$f(N(t)) = f(0) + \int_0^t [f(N(s-) + 1) - f(N(s-))] dN(s). \tag{11.6.4}$$

Setting $N(t) = M(t) + \int_0^t \lambda(N(u)) du$ with $M(t)$ a martingale, we have

$$\begin{aligned} &f(N(t)) - f(0) - \int_0^t \lambda(N(s)) [f(N(s) + 1) - f(N(s))] ds \\ &= \int_0^t [f(N(s-) + 1) - f(N(s-))] dM(s). \end{aligned}$$

Introducing the generator

$$\mathbb{A} f(n) = \lambda(n) [f(n+1) - f(n)]$$

and using the fact that $M(s)$ is a martingale, it follows that

$$f(N(t)) - f(N(0)) - \int_0^t \mathbb{A} f(N(s)) ds \tag{11.6.5}$$

is a martingale and, in particular,

$$\mathbb{E}[f(N(t))] = \mathbb{E}[f(N(0))] + \int_0^t \mathbb{E}[\mathbb{A} f(N(s))] ds. \tag{11.6.6}$$

Consider the index function $f(N) = \chi_n(N) = 1$ if $N = n$ and zero otherwise. Then $\mathbb{E}[f(N(t))] = \mathbb{P}[N(t) = n]$ and

$$\mathbb{P}[N(t)=n]=\mathbb{P}[N(0)=n]+\int_0^t\left[\lambda(n-1)\mathbb{P}[N(s)=n-1]-\lambda(n)\mathbb{P}[N(s)=n]\right].$$

Differentiating both sides with respect to t and setting $p(n,t)=\mathbb{P}[N(t)=n]$ yields the forward master equation for the counting process

$$\frac{dp(n,t)}{dt}=[\lambda(n-1)p(n-1,t)-\lambda(n)p(n,t)]. \tag{11.6.7}$$

11.6.2 Chemical Reactions and Counting Processes

Now consider the simple single-step reaction

$$A+B\xrightarrow{\kappa}C.$$

Let $\mathbf{X}(t)=(X_A(t),X_B(t),X_C(t))$ be the state of the stochastic process at time t with $X_i(t)$ the number of molecules of chemical species i at time t. From simple bookkeeping, we can write

$$\mathbf{X}(t)=\mathbf{X}(0)+R(t)\begin{pmatrix}-1\\-1\\1\end{pmatrix}, \tag{11.6.8}$$

where $R(t)$ is the number of reactions that has occurred by time t, $\mathbf{X}(0)$ is the initial state, and the constant vector specifies the stochiometric coefficients. We will assume that the probability of a reaction in an infinitesimal interval $(t,t+\Delta t]$ is

$$\mathbb{P}[\text{reaction occurs in } (t,t+\Delta t]|\mathscr{F}_t]\approx\kappa X_A(t)X_B(t)\Delta t, \tag{11.6.9}$$

where $\mathscr{F}_t$ is the information obtained by observing the stochastic process up to time t. Equation (11.6.9) is consistent with the law of mass action in the deterministic limit. The basic idea of the time-change representation is that the number of reactions can be expressed in terms of a unit rate Poisson process according to

$$R(t)=Y\left(\int_0^t\kappa X_A(s)X_B(s)ds\right). \tag{11.6.10}$$

In order to establish that this is consistent with equation (11.6.9), set $\lambda(\mathbf{X}(t))=\kappa X_A(t)X_B(t)$ and note that the probability a reaction occurs in a small time interval $(t,t+\Delta t]$ is

$$\begin{aligned}\mathbb{P}[R(t+\Delta t)>R(t)|\mathscr{F}_t]&=1-\mathbb{P}[R(t+\Delta t)=R(t)|\mathscr{F}_t]\\&=1-\mathbb{P}\left[Y\left(\int_0^{t+\Delta t}\lambda(\mathbf{X}(s))ds\right)-Y\left(\int_0^t\lambda(\mathbf{X}(s))ds\right)=0\right]\\&=1-\mathrm{e}^{-\lambda(\mathbf{X}(t))\Delta t}\approx\lambda(\mathbf{X}(t))\Delta t.\end{aligned}$$

We have used the fact that $Y\left(\int_0^t \lambda(\mathbf{X}(s))ds\right)$ and $\mathbf{X}(t)$ are part of the information in $\mathscr{F}_t$, that is, they are $\mathscr{F}_t$-measurable. Similarly, for a general set of chemical reactions involving N chemical species $i = 1,\ldots,N$ and K single-step reactions with stochiometric vectors $\mathbf{S}_a \in \mathbb{Z}^N$ and propensity functions λ_a, $a = 1,\ldots,K$ (see Sect. 6.3), we have

$$\mathbf{X}(t) = \mathbf{X}(0) + \sum_a \mathbf{S}_a R_a(t), \tag{11.6.11}$$

where $\mathbf{X}(t) = (X_1(t),\ldots,X_N(t))$ and $R_a(t)$ is the number of occurrences of reaction a up to time t. Moreover,

$$R_a(t) = Y_a\left(\int_0^t \lambda_a(\mathbf{X}(s))ds\right), \tag{11.6.12}$$

where the Y_a are independent unit rate Poisson processes. Following along similar lines to our analysis of a simple counting process, it can be shown that the counting processes $(R_1,\ldots,R_K)$ are solutions of the stochastic process defined by equations (11.6.11) and (11.6.12) if and only if they are solutions of a corresponding martingale problem with respect to the intensities λ_a: there exists a filtration $\{\mathscr{F}_t\}$ to which the R_a are adapted and

$$M_{a,k}(t) \equiv R_a(t \wedge \tau_k) - \int_0^{t\wedge\tau_k} \lambda_a(\mathbf{X}(s))ds$$

is a $\{\mathscr{F}_t\}$-martingale.

The martingale problem can be used to derive the master equation for a chemical reaction network. For simplicity, suppose that $\tau_\infty = \infty$ and $\mathbb{E}[R_a(t)] < \infty$. (This condition can be relaxed [9].) It follows that

$$M_a(t) = R_a(t) - \int_0^t \lambda_a(\mathbf{X}(s))ds$$

is a martingale so that if $R_a(0) = 0$ for all a, then

$$\mathbb{E}[R_a(t)] = \mathbb{E}\left[\int_0^t \lambda_a(\mathbf{X}(s))ds\right].$$

Introduce an arbitrary, bounded function f on $\mathbb{Z}^N$. Generalizing equation (11.6.4), we have

$$f(\mathbf{X}(t)) = f(\mathbf{X}(0)) + \sum_a \int_0^t [f(\mathbf{X}(s-) + \mathbf{S}_a) - f(\mathbf{X}(s-))]\, dR_a(s),$$

which can be rewritten as

$$f(\mathbf{X}(t)) - f(\mathbf{X}(0)) - \int_0^t \sum_a \lambda_a(\mathbf{X}(s)) \left[f(\mathbf{X}(s) + \mathbf{S}_a) - f(\mathbf{X}(s)) \right] ds$$
$$= \sum_a \int_0^t \left[f(\mathbf{X}(s-) + \mathbf{S}_a) - f(\mathbf{X}(s-)) \right] dM_a(s).$$

Introducing the generator (of the master equation)

$$\mathbb{A}f(x) = \sum_a \lambda_a(\mathbf{x}) [f(\mathbf{x} + \mathbf{S}_a) - f(\mathbf{x})]$$

and using the fact that $M_a(s)$ is a martingale, it follows that

$$f(\mathbf{X}(t)) - f(\mathbf{X}(0)) - \int_0^t \mathbb{A}f(\mathbf{X}(s)) ds \tag{11.6.13}$$

is a martingale and, in particular,

$$\mathbb{E}[f(\mathbf{X}(t))] = \mathbb{E}[f(\mathbf{X}(0))] + \int_0^t \mathbb{E}[\mathbb{A}f(\mathbf{X}(s))] ds. \tag{11.6.14}$$

Consider the index function $f(\mathbf{x}) = \chi_{\mathbf{y}}(\mathbf{x}) = 1$ if $\mathbf{y} = \mathbf{x}$ and zero otherwise. Then $\mathbb{E}[f(\mathbf{X}(t))] = \mathbb{P}[\mathbf{X}(t) = \mathbf{y}]$ and

$$\mathbb{P}[\mathbf{X}(t) = \mathbf{y}] = \mathbb{P}[\mathbf{X}(0) = \mathbf{y}]$$
$$+ \int_0^t \sum_a (\lambda_a(\mathbf{y} - \mathbf{S}_a) \mathbb{P}[\mathbf{X}(s) = \mathbf{y} - \mathbf{S}_a] - \lambda_a(\mathbf{y}) \mathbb{P}[\mathbf{X}(s) = \mathbf{y}]) ds.$$

Differentiating both sides with respect to t and setting $p(\mathbf{y}, t) = \mathbb{P}[\mathbf{X}(t) = \mathbf{y}]$ yields the forward master equation

$$\frac{dp(\mathbf{y}, t)}{dt} = \sum_a \left[\lambda_a(\mathbf{y} - \mathbf{S}_a) p(\mathbf{y} - \mathbf{S}_a, t) - \lambda_a(\mathbf{y}) p(\mathbf{y}, t) \right]. \tag{11.6.15}$$

Having shown how to reformulate chemical reaction networks in terms of counting processes and martingales, we briefly highlight some of the applications of the latter [9]:

1. The martingale properties of the counting process R_a provide an effective method for evaluating moments of the chemical processes $X_i(t)$.
2. One can obtain rigorous asymptotic estimates for convergence to the deterministic rate equations in the large N limit, where N is the system size (number of molecules). One can also carry out a rigorous slow/fast decomposition in the case of multiple time scales.

3. The concept of the generator $\mathbb{A}$ and its associated martingales can be extended to the case of continuous stochastic processes based on a Langevin equation. Indeed, carrying out a system-size expansion of the master equation generator, leads to the second-order differential operator of the FPE
4. Equations (11.6.11) and (11.6.12) provide the basis for SSAs such as Gillespie's direct method [217] (see Sect. 6.8).

References

1. Acar, M., Mettetal, J.T., van Oudenaarden, A.: Stochastic switching as a survival strategy in fluctuating environments. Nat. Genet. **40**, 471–475 (2008)
2. Aghababaie, Y., Menon, G.I., Plischke, M.: Universal properties of interacting Brownian motors. Phys. Rev. E **59**, 2578–2586 (1999)
3. Alberts, B., Johnson, A., Lewis, J., Raff, M., Walter, K.R.: Molecular Biology of the Cell, 5th edn. Garland, New York (2008)
4. Allen, L.J.S.: An Introduction to Stochastic Processes with Applications to Biology, 2nd edn. Chapman and Hall/CRC, Boca Raton (2010)
5. Alon, U.: An Introduction to Systems Biology: Design Principles of Biological Circuits. Chapman and Hall/CRC, Boca Raton (2007)
6. Alt, W.: Biased random walk model for chemotaxis and related diffusion approximation. J. Math. Biol. **9**, 147–177 (1980)
7. Altschuler, S.J., Angenent, S.B., Wang, Y., Wu, L.F.: On the spontaneous emergence of cell polarity. Nature **454**, 886–890 (2008)
8. Anderson, D.F.: A modified next reaction method for simulating chemical systems with time dependent propensities and delays. J. Chem. Phys. **127**, 214107 (2007)
9. Anderson, D.F., Kurtz, T.G.: Continuous time Markov chain models for chemical reaction networks. In: Design and Analysis of Biomolecular Circuits, pp. 3–42. Springer, New York (2011)
10. Andrews, S., Bray, D.: Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. Phys. Biol. **1**, 137–151 (2004)
11. Antal, T., Krapivsky, P.L., Redner, S., Mailman, M., Chakraborty, B.: Dynamics of an idealized model of microtubule growth and catastrophe. Phys. Rev. E **76**, 041907 (2007)
12. Arimura, N., Kaibuchi, K.: Neuronal polarity: from extracellular signals to intracellular mechanisms. Nat. Rev. Neurosci. **8**, 194–205 (2007)
13. Arkin, A., Ross, J., McAdams, H.H.: Stochastic kinetic analysis of developmental pathway bifurcation in phage infected escherichia coli cells. Genetics **149**, 1633–1648 (1998)
14. Assmann, M.-A., Lenz, P.: Characterization of bidirectional molecular motor-assisted transport models. Phys. Biol. **10**, 016003 (2013)
15. Atanasova, K.T., Burgo, A., Galli, T., Holcman, D.: Quantifying neurite growth mediated by interactions among secretory vesicles, microtubules, and actin networks. Biophys. J. **96**, 840–857 (2009)
16. Athreya, K.B., Ney, P.E.: Branching Processes. Springer, Berlin (1972)
17. Axelrod, D., Koppel, D.E., Schlessinger, J., Elson, E., Webb, W.W.: Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. Biophys. J. **16**, 1055–1069 (1976)

P.C. Bressloff, *Stochastic Processes in Cell Biology*, Interdisciplinary Applied Mathematics 41, DOI 10.1007/978-3-319-08488-6

18. Bai, L., Santangelo, T.J., Wang, M.D.: Single-molecule analysis of RNA polymerase transcription. Annu. Rev. Biophys. Biomol. Struct. **35**, 343–360 (2006)
19. Baras, F., Mansour, M., Malek, M., Pearson, J.E.: Microscopic simulation of chemical bistability in homogeneous systems. J. Chem. Phys. **105**, 8257–8261 (1996)
20. Barkai, N., Leibler, S.: Robustness in simple biochemical networks. Nature **387**, 913–917 (1997)
21. Barkai, N., Leibler, S.: Circadian clocks limited by noise. Nature **403**, 267–268 (2000)
22. Barkai, E., Silbey, R.: Theory of single file diffusion in a force field. Phys. Rev. Lett. **102**, 050602 (2009)
23. Barnhart, E.L., Lee, K.C., Keren, K., Mogilner, A., Theriot, J.A.: An adhesion-dependent switch between mechanisms that determine motile cell shape. PLoS Biol. **9**, e1001059 (2011)
24. Barrio, B.A., Varea, C., Aragon, J.L., Maini, P.K.: A two-dimensional numerical study of spatial pattern formation in interacting Turing systems. Bull. Math. Biol. **61**, 483 (1999)
25. Basu, A., Chowdhury, D.: Traffic of interacting ribosomes: effects of single-machine mechanochemistry on protein synthesis. Phys. Rev. E **75**, 021902 (2007)
26. Becalska, A.N., Gavis, E.R.: Lighting up mRNA localization in Drosophila oogenesis. Development **136**, 2493–2503 (2009)
27. Becskei, A., Mattaj, I.W.: Quantitative models of nuclear transport. Curr. Opin. Cell Biol. **17**, 27–34 (2005)
28. Bell, J.W.: Searching Behavior, the Behavioral Ecology of Finding Resources. Chapman and Hall, London (1991)
29. Benichou, O., Voituriez, R.: Narrow escape time problem: time needed for a particle to exit a confining domain through a small window. Phys. Rev. Lett. **100**, 168105 (2008)
30. Benichou, O., Coppey, M., Moreau, M., Suet, P., Voituriez, R.: A stochastic model for intermittent search strategies. J. Phys. Condens. Matter **17**, S4275–S4286 (2005)
31. Benichou, O., Loverdo, C., Moreau, M., Voituriez, R.: A minimal model of intermittent search in dimension two. J. Phys. A **19** 065141 (2007)
32. Benichou, O., Loverdo, C., Moreau, M., Voituriez, R.: Intermittent search strategies. Rev. Mod. Phys. **83**, 81 (2011)
33. Berezhkovskii, A.M., Bezrukov, S.M.: Channel-facilitated membrane transport: constructive role of the particle attraction to the channel pore. Chem. Phys. **319**, 342–349 (2005)
34. Berezhkovskii, A.M., Pusovoit, M.A., Bezrukov, S.M.: Channel-facilitated membrane transport: transit probability and interaction with the channel. J. Chem. Phys. **116**, 9952–9956 (2002)
35. Berezhkovskii, A.M., Pusovoit, M.A., Bezrukov, S.M.: Channel-facilitated membrane transport: average lifetimes in the channel. J. Chem. Phys. **119**, 3943–3951 (2003)
36. Berg, O.G.: A model for the statistical fluctuations of protein numbers in a microbial population. J. Theor. Biol. **4**, 587–603 (1978)
37. Berg, H.C.: Random Walks in Biology. Princeton University Press, Princeton (1983)
38. Berg, H.C.: Motile behavior of bacteria. Phys. Today **53**, 24–28 (2000)
39. Berg, H.C., Brown, D.A.: Chemotaxis in Escherichia coli analyzed by three-dimensional tracking. Nature **239**, 500–504 (1972)
40. Berg, H.C., Purcell, E.M.: Physics of chemoreception. Biophys. J. **20**, 93–219 (1977)
41. Berg, O.G., von Hippel, P.H.: Diffusion-controlled macromolecular interactions. Annu. Rev. Biophys. Biophys. Chem. **14**, 131–160 (1985)
42. Berg, O.G., Winter, R.B., von Hippel, P.H.: Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. Biochemistry **20**, 6929–6948 (1981)
43. Bhalla, U.S.: Signaling in small subcellular volumes I. stochastic and diffusion effects on individual pathways. Biophys. J. **87**, 733–744 (2004)
44. Bialek, W.: Biophysics. Princeton University Press, Princeton (2012)
45. Bialek, W., Setayeshgar, S.: Physical limits to biochemical signaling. Proc. Natl. Acad. Sci. USA **102**, 10040–10045 (2005)
46. Bialek, W., Setayeshgar, S: Cooperativity, sensitivity and noise in biochemical signaling. Phys. Rev. Lett. **100**, 258101 (2008)

47. Biancalani, T., Fanelli, D., Di Patti, F.: Stochastic turing patterns in the Brusselator model. Phys. Rev. E **81**, 046215 (2010)
48. Biancalani, T., Galla, T., McKane, A.J.: Stochastic waves in a Brusselator model with non-local interaction. Phys. Rev. E **84**, 026201 (2011)
49. Bickel, T., Bruinsma, R.: The nuclear pore complex mystery and anomalous diffusion in reversible gels. Biophys. J. **83**, 3079–3987 (2002)
50. Bicout, D.J.: Green's functions and first passage time distributions for dynamic instability of microtubules. Phys. Rev. E **56**, 6656–6667 (1997)
51. Binder, B., Goede, A., Berndt, N., Holzhutter, H.-G.: A conceptual mathematical model of the dynamic self-organisation of distinct cellular organelles. PLoS One **4**, e8295 (2009)
52. Blythe, R.A., Evans, M.R.: Nonequilibrium steady states of matrix-product form: a solver's guide. J. Phys. A **40**, R333–R441 (2007)
53. Boal, D.: Mechanics of the Cell, 2nd edn. Cambridge University Press, Cambridge (2010)
54. Boettiger, D.: Mechanical control of integrin-mediated adhesion and signaling. Curr. Opin. Cell Biol. **24**, 592–599 (2012)
55. Boland, R.P., Galla, T., McKane, A.J.: How limit cycles and quasi-cycles are related in systems with intrinsic noise. J. Stat. Mech. Theory Exp. **P09001**, 1–27 (2008)
56. Boland, R.P., Galla, T., McKane, A.J.: Limit cycles, complex floquet multipliers, and intrinsic noise. Phys. Rev. E **79**, 051131 (2009)
57. Bollenbach, T., Kruse, K., Pantazis, P., Gonzalez-Gaitan, M., Julicher, F.: Robust formation of morphogen gradients. Phys. Rev. Lett. **94**, 018103 (2005)
58. Bouchaud, J.P., Georges, A.: Anomalous diffusion in disordered media—statistical mechanisms, models and physical applications. Phys. Rep. **195**, 127–293 (1990)
59. Bouzigues, C., Morel, M., Triller, A., Dahan, M.: Asymmetric redistribution of GABA receptors during GABA gradient sensing by nerve growth cones analyzed by single quantum dot imaging. Proc. Natl. Acad. Sci. USA **104**, 11251–11256 (2007)
60. Brandenburg, B.: Virus trafficking-learning from single-virus tracking. Nat. Rev. Microbiol. **5**, 197–208 (2007)
61. Bray, D.: Cell Movements, 2nd edn. Garland, New York (2001)
62. Bray, D., Levin, M.D., Morton-Firth, C.J.: Receptor clustering as a cellular mechanism to control sensitivity. Nature **393**, 85–88 (1998)
63. Bredt, D.S., Nicoll, R.A.: AMPA receptor trafficking at excitatory synapses. Neuron **40**, 361–379 (2003)
64. Bressloff, P.C.: A stochastic model of intraflagellar transport. Phys. Rev. E **73**, 061916 (2006)
65. Bressloff, P.C.: Two-pool model of cooperative vesicular transport. Phys. Rev. E **86**, 031911 (2012)
66. Bressloff, P.C.: Waves in Neural Media: From Single Neurons to Neural Fields. Springer, New York (2014)
67. Bressloff, P.C., Earnshaw, B.A.: Diffusion-trapping model of receptor trafficking in dendrites. Phys. Rev. E **75**, 041916 (2007)
68. Bressloff, P.C., Earnshaw, B.A.: A dynamical corral model of protein trafficking in spines. Biophys. J. **96**, 1786–1802 (2009)
69. Bressloff, P.C., Newby, J.M.: Directed intermittent search for hidden targets. New J. Phys. **11**, 023033 (2009)
70. Bressloff, P.C., Newby, J.M.: Quasi-steady state analysis of motor-driven transport on a two-dimensional microtubular network. Phys. Rev. E **83**, 061139 (2011)
71. Bressloff, P.C., Newby, J.M.: Filling of a Poisson trap by a population of random intermittent searchers. Phys. Rev. E **85**, 031909 (2012)
72. Bressloff, P.C., Newby, J.M.: Stochastic models of intracellular transport. Rev. Mod. Phys. **85**, 135–196 (2013)
73. Bressloff, P.C., Newby, J.M.: Path-integrals and large deviations in stochastic hybrid systems. Phys. Rev. E **89**, 042701 (2014)
74. Bressloff, P.C., Newby, J.M.: Stochastic hybrid model of spontaneous dendritic NMDA spikes. Phys. Biol. **11**, 016006 (2014)

75. Bressloff, P.C., Cowan, J.D., Golubitsky, M., Thomas, P.J., Wiener, M.: Geometric visual hallucinations, euclidean symmetry and the functional architecture of striate cortex. Phil. Trans. Roy. Soc. Lond. B **356**, 299–330 (2001)
76. Bressloff, P.C., Earnshaw B.A., Ward, M.J.: Diffusion of protein receptors on a cylindrical dendritic membrane with partially absorbing traps. SIAM J. Appl. Math. **68**, 1223–1246 (2008)
77. Bringuier, E.: Kinetic theory of inhomogeneous diffusion. Phys. A **388**, 2588–2599 (2009)
78. Brown, G.C., Kholodenko, B.N.: Spatial gradients of cellular phospho-proteins. FEBS Lett. **457**, 452–454 (1999)
79. Brown, F.L.H., Leitner, D.M., McCammon, J.A., Wilson, K.R.: Lateral diffusion of membrane proteins in the presence of static and dynamics corrals: suggestions for appropriate variables. Biophys. J. **78**, 2257–2269 (2000)
80. Brust-Mascher, I., Civelekoglu-Scholey, G., Kwon, M., Mogilner, A., Scholey, J.M.: Model for anaphase B: role of three mitotic motors in a switch from poleward flux to spindle elongation. Proc. Natl. Acad. Sci. USA **101**, 15938–15943 (2004)
81. Buckwar, E., Riedler, M.G.: An exact stochastic hybrid model of excitable membranes including spatio-temporal evolution. J. Math. Biol. **63**, 1051–1093 (2011)
82. Burada, P.S., Schmid, G., Reguera, D., Rubi, J.M., Hanggi, P.: Biased diffusion in confined media: test of the Fick-Jacobs approximation and validity criteria. Phys. Rev. E **75**, 051111 (2007)
83. Burada, P.S., Hanggi, P., Marchesoni, F., Schmid, G., Talkner, P.: Diffusion in confined geometries. Chem. Phys. Chem. **10**, 45–54 (2009)
84. Burbank, K.S., Mitchison, T.J., Fisher, D.S.: Slide-and-cluster models for spindle assembly. Curr. Biol. **17**, 1373–1383 (2007)
85. Burlakov, V.M., Emptage, N., Goriely, A., Bressloff, P.C.: Phys. Rev. Lett. **108**, 028101 (2012)
86. Burov, S., Jeon, J.H., Metzler, R., Barkai, E.: Single particle tracking in systems showing anomalous diffusion: the role of weak ergodicity breaking. Phys. Chem. Chem. Phys. **13**, 1800–1812 (2011)
87. Butler, T.C., Goldenfeld, N.: Robust ecological pattern formation induced by demographic noise. Phys. Rev. E **80**, 030902(R) (2009)
88. Butler, T.C., Goldenfeld, N.: Fluctuation-driven turing patterns. Phys. Rev. E **84**, 011112 (2011)
89. Cai, L., Friedman, N., Xies, X.S.: Stochastic protein expression in individual cells at the single molecule level. Nature **440**, 358–362 (2006)
90. Camalet, S., Duke, T., Julicher, F., Prost, J.: Auditory sensitivity provided by self-tuned critical oscillations of hair cells. Proc. Natl. Acad. Sci. USA **97**, 3183–3188 (2000)
91. Cameron, L.A., Giardini, P.A., Soo, F.S., Theriot, J.A.: Secrets of actin-based motility revealed by a bacterial pathogen. Nat. Rev. Mol. Cell Biol. **1**, 110–119 (2000)
92. Campas, O., Leduc, C., Bassereau, P., Joanny, J.- F., Prost, J.: Collective oscillations of processive molecular motors. Biophys. Rev. Lett. **4**, 163–178 (2009)
93. Cao, Y., Gillespie, D.T., Petzold, L.R.: Efficient step size selection for the tau-leaping simulation method. J. Chem. Phys. **124**, 044109 (2006)
94. Capas, O., Sens, P.: Chromosome oscillations in mitosis. Phys. Rev. Lett. **97**, 128102 (2006)
95. Carazo-Salas, R.E., et al.: Ran-GTP coordinates regulation of microtubule nucleation and dynamics during mitotic-spindle assembly. Nat. Cell Biol. **3**, 228–234 (2001)
96. Carlsson, A.E.: Growth of branched actin networks against obstacles. Biophys. J. **81**, 1907–1923 (2001)
97. Cassimeris, L., Rieder, C., Salmon, E.: Microtubule assembly and kinetochore directional instability in vertebrate monopolar spindles: implications for the mechanism of chromosome congression. J. Cell Sci. **107**, 285–297 (1994)
98. Caudron, M., Bunt, G., Bastiaens, P., Karsenti, E.: Spatial coordination of spindle assembly by chromosome-mediated signaling gradients. Science **309**, 1373–1376 (2005)

99. Celton-Morizur, S., Racine, V., Sibarita, J.B., Paoletti, A.: Pom1 kinase links division plane position to cell polarity by regulating Mid1p cortical distribution. J. Cell Sci. **119**, 4710–4718 (2006)
100. Centres, P.M., Bustingorry, S.: Effective Edwards-Wilkinson equation for single-file diffusion. Phys. Rev. E **81**, 061101 (2010)
101. Chan, C.E., Odde, D.J.: Traction dynamics of filopodia on compliant substrates. Science **322**, 1687–1691 (2008)
102. Chandler, D.: Introduction to modern statistical mechanics. Oxford University Press, Oxford (1987)
103. Cheng, H., Lederer, W.J.: Calcium sparks. Physiol. Rev. **88**, 1491–1545 (2008)
104. Cheng, H., Lederer, M.R., Lederer, W.J., Cannell, M.B.: Calcium sparks and waves in cardiac myocytes. Am. J. Physiol. **270**, C148–C159 (1996)
105. Cheong, R., Rhee, A., Wang, C.J., Nemenman, I., Levchenko, A.: Information transduction capacity of noisy biochemical signaling networks. Science **334**, 354–357 (2011)
106. Cherstvy, A.G., Kolomeisky, A.B., Kornyshev, A.A.: Protein–DNA interactions: reaching and recognizing the targets. J. Phys. Chem. **112**, 4741–4750 (2008)
107. Cheviakov, A.F., Ward, M.J., Straube, R.: An asymptotic analysis of the mean first passage time for narrow escape problems: Part II: the sphere. SIAM J. Multiscale Model. Simul. **8**, 836–870 (2010)
108. Chipot, M., Hastings, S., Kinderlehrer, D.: Transport in a molecular motor system. Math. Model. Numer. Anal. **38**, 1011–1034 (2004)
109. Choe, Y., Magnasco, M.O., Hudspeth, A.J.: A model for amplification of hair-bundle motion by cyclical binding of Ca^{2+} to mechanoelectrical transduction channels. Proc. Natl. Acad. Sci. USA **95**, 15321–15326 (1998)
110. Choi, C.K., Vicente-Manzanares, M., Zareno, J., Whitmore, L.A., Mogilner, A., Horwitz, A.R.: Actin and α-actinin orchestrate the assembly and maturation of nascent adhesions in a myosin II motor-independent manner. Nat. Cell Biol. **10**, 1039–1050 (2008)
111. Choquet, D., Triller, A.: The role of receptor diffusion in the organization of the postsynaptic membrane. Nat. Rev. Neurosci. **4**, 251–265 (2003)
112. Chou, T.: Kinetics and thermodynamics across single-file pores: solute permeability and rectified osmosis. J. Chem. Phys. **110**, 606–615 (1999)
113. Chou, T., Lakatos, G.: Totally asymmetric exclusion processes with particles of arbitrary size. J. Phys. A **36**, 2027–2041 (2003)
114. Chou, T., Lakatos, G.: Clustered bottlenecks in mRNA translation and protein synthesis. Phys. Rev. Lett. **93**, 198101 (2004)
115. Chou, T., Mallick, K., Zia, R.K.P.: Non-equilibrium statistical mechanics: from a paradigmatic model to biological transport. Rep. Prog. Phys. **74**, 116601 (41 pp) (2011)
116. Chow, C.C., White, J.A.: Spontaneous action potentials due to channel fluctuations. Biophys. J. **71**, 3013–3021 (1996)
117. Chuang, J., Kantor, Y., Kardar, M.: Anomalous dynamics of translocation. Phys. Rev. E **65**, 011802 (2002)
118. Ciandrini, L., Stansfield, I., Romano, M.C.: Role of the particle's stepping cycle in an asymmetric exclusion process: a model of mRNA translation. Phys. Rev. E **81**, 051904 (2010)
119. Civelekoglu-Scholey, G., Scholey, J.M.: Mitotic force generators and chromosome segregation. Cell. Mol. Life Sci. **67**, 2231–2250 (2010)
120. Civelekoglu-Scholey, G., Sharp, D.J., Mogilner, A., Scholey, J.M.: Model of chromosome motility in Drosophila embryos: adaptation of a general mechanism for rapid mitosis. Biophys. J. **90**, 3966–3982 (2006)
121. Clausznitzer, D., Linder, B., Julicher, F., Martin, P.: Two-state approach to stochastic hair bundle dynamics. Phys. Rev. E **77**, 041901 (2008)
122. Cohen, N., Boyle, J.J.: Swimming at low Reynolds number: a beginners guide to undulatory locomotion. Contemp. Phys. **51**, 103–123 (2009)
123. Collinridge, G.L., Isaac, J.T.R., Wang, Y.T.: Receptor trafficking and synaptic plasticity. Nat. Rev. Neurosci. **5**, 952–962 (2004)

124. Collins, F.C., Kimball, G.E.: Diffusion-controlled reaction rates. J. Colloid Sci. **4**, 425–439 (1949)
125. Coombs, D., Goldstein, B.: T cell activation: kinetic proofreading, serial engagement and cell adhesion. J. Comput. Appl. Math. **184**, 121–139 (2005)
126. Coombs, D., Straube, R., Ward, M.J.: Diffusion on a sphere with localized traps: mean first passage time, eigenvalue asymptotics, and fekete points. SIAM J. Appl. Math. **70**, 302–332 (2009)
127. Coppey, M., Benichou, O., Voituriez, R., Moreau, M.: Kinetics of target site localization of a protein DNA: a stochastic approach. Biophys. J. **87**, 1640–1649 (2004)
128. Coulon, A., Chow, C.C., Singer, R.H., Larson, D.R.: Eukaryotic transcriptional dynamics: from single molecules to cell populations. Nat. Rev. Genet. **14**, 572–584 (2013)
129. Cox, J.S., Chapman, R.E., Walter, P.: The unfolded protein response coordinates the production of endoplasmic reticulum protein and endoplasmic reticulum membrane. Mol. Biol. Cell **8**, 1805–1814 (1997)
130. Cronshaw, J.M., Matunis, M.J.: The nuclear pore complex: disease associations and functional correlations. Trends Endocrinol. Metab. **15**, 34–39 (2004)
131. Cytrynbaum, E.N., Scholey, J.M., Mogilner, A.: A force balance model of early spindle pole separation on *drosophila* embryos. Biophys. J. **84**, 757–769 (2003)
132. Czondora, K., Mondina, M., Garciaa, M., Heinec, M., Frischknechtc, R., Choquet, D., Sibaritaa, J.B., Thoumine, O.R.: A unified quantitative model of AMPA receptor trafficking at synapses. Proc. Natl. Acad. Sci. USA **109**, 3522–3527 (2012)
133. Damm, E.M., Pelkmans, L.: Systems biology of virus entry in mammalian cells. Cell Microbiol. **8**, 1219–1227 (2006)
134. Danuser, G., Allard, J., Mogilner, A.: Mathematical modeling of eukaryotic cell migration: insights beyond experiments. Annu. Rev. Cell Dev. Biol. **29**, 501–528 (2013)
135. Davis, M.H.A.: Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. Journal of the Royal Society, Series B (Methodological), **46**, 353–388 (1984)
136. Delorme, V., Machacek, M., DerMardirossian, C., Anderson, K.L., Wittmann, T., Hanein, D., Waterman-Storer, C., Danuser, G., Bokoch, G.M.: Cofilin activity downstream of Pak1 regulates cell protrusion efficiency by organizing lamellipodium and lamella actin networks. Dev. Cell. **13**, 646–662 (2007)
137. Derrida, B.: Velocity and diffusion constant of a periodic one-dimensional hopping model. J. Stat. Phys. **31**, 433–450 (1983)
138. Dickinson, R.B.: Models for actin polymerization. J. Math. Biol. **58**, 81–103 (2009)
139. Dickinson, R.B., Purich, D.L.: Clamped-filament elongation model for actin-based motors. Biophys. J. **82**, 605–617 (2002)
140. Dickinson, R.B., Tranquillo, R.T.: Transport equations and indices for random and biased cell migration based on single cell properties. SIAM J. Appl. Math. **55**, 1419–1454 (1995)
141. DiMilla, P.A., Barbee, K., Lauffenburger, D.A.: Mathematical model for the effects of adhesion and mechanics on cell migration speed. Biophys. J. **60**, 15–37 (1991)
142. Dinh, A.T., Pangarkar, C., Theofanous, T., Mitragotri, S.: Understanding intracellular transport processes pertinent to synthetic gene delivery via stochastic simulations and sensitivity analysis. Biophys. J. **92**, 831–846 (2007)
143. Dix, J.A., Verkman, A.S.: Crowding effects on diffusion in solutions and cells. Annu. Rev. Biophys. **37**, 247–263 (2008)
144. Dmitrieff, S., Sens, P.: Cooperative protein transport in cellular organelles. Phys. Rev. E **83**, 041923 (2011)
145. Doering, C.R. , Sargsyan, K.V., Sander, L.M., Vanden-Eijnden, E.: Asymptotics of rare events in birth-death processes bypassing the exact solutions. J. Phys. Condens. Matter **19**, 065145 (2007)
146. Dogterom, M., Leibler, S.: Physical aspects of the growth and regulation of microtubule structures. Phys. Rev. Lett. **70**, 1347–1350 (1993)

147. Doi, M.: Second quantization representation for classical many-particle systems. J. Phys. A **9**, 1465–1477 (1976)
148. Doi, M.: Stochastic theory of diffusion controlled reactions. J. Phys. A **9**, 1479–1495 (1976)
149. Dong, J., Schmittmann, B., Zia., R.K.P.: Inhomogeneous exclusion processes and protein synthesis. Phys. Rev. E **76**, 051113 (2007)
150. Driver, J.W., Rodgers, A.R., Jamison, D.K., Das, R.K., Kolomeisky, A.B., Diehl, M.R.: Coupling between motor proteins determines dynamic behavior of motor protein assemblies. Phys. Chem. Chem. Phys. **12**, 10398–10405 (2010)
151. Duke T.A.J., Bray, D.: Heightened sensitivity of a lattice of membrane of receptors. Proc. Natl. Acad. Sci. USA **96**, 10104–10108 (1999)
152. Dykman, M.I., Mori, E., Ross, J., Hunt, P.M.: Large fluctuations and optimal paths in chemical kinetics. J. Chem. Phys. A **100**, 5735–5750 (1994)
153. Dynes, J., Steward, O.: Dynamics of bidirectional transport of ARC mRNA in neuronal dendrites. J. Comput. Neurol. **500**, 433–447 (2007)
154. Earnshaw, B.A., Bressloff, P.C.: A biophysical model of AMPA receptor trafficking and its regulation during LTP/LTD. J. Neurosci. **26**, 12362–12373 (2006)
155. Earnshaw, B.A., Bressloff, P.C.: Modeling the role of lateral membrane diffusion on AMPA receptor trafficking along a spiny dendrite. J. Comput. Neurosci. **25**, 366–389 (2008)
156. Edelstein-Keshet, L., Ermentrout, G.B.: Models for the length distribution of actin filaments: simple polymerization and fragmentation. Bull. Math. Biol. **60**, 449–475 (1998)
157. Edelstein-Keshet, L., Ermentrout, G.B.: Models for spatial polymerization dynamics of rod-like polymers. J. Math. Biol. **40**, 64–96 (2000)
158. Eguíluz, V.M., Ospeck, M., Choe, Y., Hudspeth, A.J., Magnasco, M.O.: Essential nonlinearities in hearing. Phys. Rev. Lett. **84**, 5232–5235 (2000)
159. Ehlers, M.D.: Reinsertion or degradation of AMPA receptors determined by activity-dependent endocytic sorting. Neuron **28**, 511–525 (2000)
160. Ehlers, M.D., Heine, M., Groc, L., Lee, M.-C., Choquet, D.: Diffusional trapping of GluR1 AMPA receptors by input-specific synaptic activity. Neuron **54**, 447–460 (2007)
161. Eldar, A., Rosin, D., Shilo, B.-Z., Barkai, N.: Self-enhanced ligand degradation underlies robustness of morphogen gradients. Dev. Cell **5**, 635–646 (2003)
162. Elf, J., Ehrenberg, M.: Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. Genome Res. **13**, 2475–2484 (2003)
163. Elgart, V., Kamenev, A.: Rare event statistics in reaction–diffusion systems. Phys. Rev. E **70**, 041106 (2004)
164. Elowitz, M.B., Levine, A.J., Siggia, E.D., Swain, P.S.: Stochastic gene expression in a single cell. Science **297**, 1183–1186 (2002)
165. Elston, T.C.: A macroscopic description of biomolecular transport. J. Math. Biol. **41**, 189–206 (2000)
166. Elston, T.C., Wang, H., Oster, G.: Energy transduction in ATP synthase. Nature **391**, 510–513 (1998)
167. England, J.L., Cardy, J.: Morphogen gradient from a noisy source. Phys. Rev. Lett. **94**, 078101 (2005)
168. Erban, R., Chapman, J.: Reactive boundary conditions for stochastic simulations of reaction-diffusion processes. Phys. Biol. **4**, 16–28 (2007)
169. Erban, R., Chapman, J.: Stochastic modelling of reaction–diffusion processes: algorithms for bimolecular reactions. Phys. Biol. **6**, 046001 (2009)
170. Erban, R., Othmer, H.: From individual to collective behavior in bacterial chemotaxis. SIAM J. Appl. Math. **65**, 361–391 (2005)
171. Erban, R., Othmer, H.: From signal transduction to spatial pattern formation in *E. coli*: a paradigm for multi-scale modeling in biology. Multiscale Model. Simul. **3**, 362–394 (2005)
172. Ermentrout, G.B., Cowan, J.: A mathematical theory of visual hallucination patterns. Biol. Cybern. **34**, 137–150 (1979)
173. Ermentrout, G.B., Terman, D.: Mathematical Foundations of Neuroscience. Springer, Berlin (2010)

174. Escudero, C., Kamanev, A.: Switching rates of multistep reactions. Phys. Rev. E **79**, 041149 (2009)
175. Evans, E.: Probing the relation between force-lifetime-and chemistry in single molecular bonds. Annu. Rev. Biophys. Biomol. Struct. **30**, 105–128 (2001)
176. Evans, M.R., Juhasz, R., Santen, L.: Shock formation in an exclusion process with creation and annihilation. Phys. Rev. E **68**, 026117 (2003)
177. Faggionato, A., Gabrielli, D, Crivellari, M.: Averaging and large deviation principles for fully-coupled piecewise deterministic Markov processes and applications to molecular motors. Markov Processes and Related Fields. **16**, 497–548 (2010)
178. Falcke, M.: On the role of stochastic channel behavior in intracellular Ca^{2+} dynamics. Biophys. J. **84**, 42–56 (2003)
179. Falcke, M.: Reading the patterns in living cells: the physics of Ca^{2+} signaling. Adv. Phys. **53**, 255–440 (2004)
180. Falcke, M., Tsimiring, L., Levine, H.: Stochastic spreading of intracellular Ca^{2+} release. Phys. Rev. E **62**, 2636–2643 (2000)
181. Fedotov, S., Al-Shami, H., Ivanov, A., Zubarev, A.: Anomalous transport and nonlinear reactions in spiny dendrites. Phys. Rev. E **82**, 041103 (2010)
182. Ferenz, N.P., Paul, R., Fagerstrom, C., Mogilner, A., Wadsworth, P.: Dynein antagonizes eg5 by crosslinking and sliding antiparallel microtubules. Curr. Biol. **19**, 1833–1838 (2009)
183. Fisher, M., Kolomeisky, A.: Simple mechanochemistry describes the dynamics of kinesin molecules. Proc. Natl. Acad. Sci. USA **98**, 7748–7753 (2001)
184. Fischer-Friedrich, E., Meacci, G., Lutkenhaus, J., Chate, H., Kruse, K.: Intra- and intercellular fluctuations in Min-protein dynamics decrease with cell length. Proc. Natl. Acad. Sci. USA **107**, 6134–6139 (2010)
185. Flyvbjerg, H., Holy, T., Leibler, S.: Stochastic dynamics of microtubules—a model for caps and catastrophes. Phys. Rev. Lett. **73**, 2372–2375 (1994)
186. Flyvbjerg, H., Holy, T., Leibler, S.: Microtubule dynamics: caps, catastrophes, and coupled hydrolysis. Phys. Rev. E **54**, 5538–5560 (1996)
187. Forger, D.B., Kim, J.K.: A mechanism for robust circadian timekeeping via stoichiometric balance. Mol. Syst. Biol. **8**, 630 (1–12) (2012)
188. Forger, D.B., Peskin, C.S.: A detailed predictive model of the mammalian circadian clock. Proc. Natl. Acad. Sci. USA **100**, 14806–14811 (2003)
189. Forger, D.B., Peskin, C.S.: Stochastic simulation of the mammalian circadian clock. Proc. Natl. Acad. Sci. USA **102**, 321–324 (2005)
190. Foulaadvand, M.E., Kolomeisky, A.B., Teymouri, H.: Asymmetric exclusion processes with disorder: effect of correlations. Phys. Rev. E **78**, 061116 (2008)
191. Fox, R.F., Lu, Y.N.: Emergent collective behavior in large numbers of globally coupled independent stochastic ion channels. Phys. Rev. E **49**, 3421–3431 (1994)
192. Franks, K.M., Bartol, T.M., Sejnowski, T.J.: An MCell model of calcium dynamics and frequency-dependence of calmodulin activation in dendritic spines. Neurocomputing **38**, 9–16 (2001)
193. Franz, B., Flegg, M.B., Chapman, S.J., Erban, R.: Multiscale reaction-diffusion algorithms: PDE-assisted Brownian dynamics. SIAM J. Appl. Math. **73**, 1224–1247 (2013)
194. Freche, D., Pannasch, U., Rouach, N., Holcman, D.: Synapse geometry and receptor dynamics modulate synaptic strength. PLoS One **6**, e25122 (2011)
195. Freidlin, M.I., Wentzell, A.D.: Random Perturbations of Dynamical Systems. Springer, New York (1998)
196. Friedman, A., Craciun, G.: A model of intracellular transport of particles in an axon. J. Math. Biol. **51**, 217–246 (2005)
197. Friedman, A., Craciun, G.: Approximate traveling waves in linear reaction-hyperbolic equations. SIAM J. Math. Anal. **38**, 741–758 (2006)
198. Friedman, A., Hu, B.: Uniform convergence for approximate traveling waves in linear reaction-hyperbolic systems. Indiana Univ. Math. J. **56**, 2133–2158 (2007)

199. Friedman, N., Cai, L., Xies, X.S.: Linking stochstic dynamics to population distribution: an analytical framework of gene expression. Phys. Rev. Lett. **97**, 168302 (2006)
200. Gamba, A., Kolokolov, I., Lebedev, V., Ortenzi, G.: Universal features of cell polarization processes. J. Stat. Mech. **2009**, Po2019 (2009)
201. Gao, Y.Q.: A simple theoretical model explains dynein's response to load. Biophys. J. **90**, 811–821 (2006)
202. Garai, A., Chowdhury, D., Ramakrishnan, T.V.: Stochastic kinetics of ribosomes: single motor properties and collective behavior. Phys. Rev. E **80**, 011908 (2009)
203. Gardel, M.L., Schneider, I.C., Aratyn-Schaus, Y., Waterman, C.M.: Mechanical integration of actin and adhesion dynamics in cell migration. Annu. Rev. Cell Dev. Biol. **26**, 315–333 (2010)
204. Gardiner, C.W.: Handbook of Stochastic Methods, 4th edn. Springer, Berlin (2009)
205. Gardner, M.K., Odde, D.: Modeling of chromosome motility during mitosis. Curr. Opin. Cell Biol. **18**, 639–647 (2006)
206. Gardner, T.S., Cantor, C.R., Collins, J.J.: Construction of a genetic toggle switch in *E. coli*. Nature **403**, 339–342 (2000)
207. Gardner, M.K., et al.: Tension-dependent regulation of microtubule dynamics at kinetochores can explain metaphase congress ion in yeast. Mol. Biol. Cell **16**, 3764–3775 (2005)
208. Gardner, M.K., Zanic, M., Gell, C., Bormuth, V., et al.: Depolymerizing kinesins Kip3 and MCAK shape cellular microtubule architecture by differential control of catastrophe. Cell **147**, 1092–1103 (2011)
209. Gardner, M.K., Zanic, M., Howard, J.: Microtubule catastrophe and rescue. Curr. Opin. Cell Biol. **25**, 14–22 (2013)
210. Gaspard, P.: The correlation time of mesoscopic chemical clocks. J. Chem. Phys. **117**, 8905–8916 (2002)
211. Gerbal, F., Chaikin, P., Rabin, Y., Prost, J.: An elastic analysis of Listeria monocytogenes propulsion. Biophys. J. **79**, 2259–2275 (2000)
212. Gerland, U., Bundschuh, R., Hwa, T.: Translocation of structured polynucleotides through nanopores. Phys. Biol. **1**, 19–25 (2004)
213. Gerrow, K., Triller, A.: Synaptic stability and plasticity in a floating world. Curr. Opin. Neurobiol. **20**, 631–639 (2010)
214. Gerstner, J.R., Yin, J.C.P.: Circadian rhythms and memory formation. Nat. Rev. Neurosci. **11**, 577–588 (2010)
215. Gibson, M., Bruck, J.: Efficient exact stochastic simulation of chemical systems with many species and many channels. J. Phys. Chem. **104**, 1876–1889 (2000)
216. Gierer, A., Meinhardt, H.: A theory of biological pattern formation. Kybernetik **12**, 30–39 (1972)
217. Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. **81**, 2340–2361 (1977)
218. Gillespie, D.T.: Approximate accelerated stochastic simulation of chemically reacting systems. J. Chem. Phys. **115**, 1716–1733 (2001)
219. Gillespie, D.T., Hellander, A., Petzold, L.R.: Perspective: stochastic algorithms for chemical kinetics. J. Chem. Phys. **138**, 170901 (2013)
220. Glansdorff, P., Prigogine, I.: Thermodynamic Theory of Structure, Stability and Fluctuations. Wiley-Interscience, Chichester (1971)
221. Glendinning, P.: Stability, Instability and Chaos: An Introduction to the Theory of Nonlinear Differential Equations. Cambridge University Press, Cambridge (1994)
222. Goldbeter, A.: A model for circadian oscillations in the Drosophila period protein (PER). Proc. Roy. Soc. Lond. B Biol. Sci. **261**, 319–324 (1995)
223. Goldstein, S.: On diffusion by discontinuous movements and the telegraph equation. Quart. J. Mech. Appl. Math. **4**, 129–156 (1951)
224. Goldwyn, J.H., Shea-Brown, E.: The what and where of adding channel noise to the Hodgkin-Huxley equations. PLoS Comput. Biol. **7**, e1002247 (2011)

225. Golubitsky, M., Stewart, I., Schaeffer, D.G.: Singularities and Groups in Bifurcation Theory II. Springer, Berlin (1988)
226. Gong, H., Guo, Y., Linstedt, A., Schwartz, R.: Discrete, continuous and stochastic models of protein sorting in the Golgi. Phys. Rev. E **81**, 011914 (2010)
227. Gonze, D., Halloy, J.: Robustness of circadian rhythms with respect to molecular noise. Proc. Natl. Acad. Sci. USA **99**, 673–678 (2002)
228. Gonze, D., Halloy, J., Gaspard, P.: Biochemical clocks and molecular noise: theoretical study of robustness factors. J. Chem. Phys. **116**, 10997–11010 (2002)
229. Gopalakrishnan, M., Govindan, B.S.: A first-passage-time theory for search and capture of chromosomes by microtubules in mitosis. Bull. Math. Biol. **73**, 2483–2506 (2011)
230. Gorman, J., Greene, E.C.: Visualizing one-dimensional diffusion of proteins along DNA. Nat. Struct. Mol. Biol. **15**, 768–774 (2008)
231. Goryachev, A.B., Pokhilko, A.V.: Dynamics of Cdc42 network embodies a turing-type mechanism of yeast cell polarity. FEBS Lett. **582**, 1437–1443 (2008)
232. Goshima, G., Scholey, J.M.: Control of mitotic spindle length. Annu. Rev. Cell Dev. Biol. **26**, 21–57 (2010)
233. Goshima, G., Wollman, R., Stuurman, N., Scholey, J.M., Vale, R.D.: Length control of the metaphase spindle. Curr. Biol. **15**, 1979–1988 (2005)
234. Govindan, B.S., Gopalakrishnan, M., Chowdhury, D.: Length control of microtubules by depolymerizing motor proteins. Europhys. Lett. **83**, 40006 (2008)
235. Graham, R., Tel, T.: On the weak-noise limit of Fokker-Planck models. J. Stat. Phys. **35**, 729 (1984)
236. Graham, R., Tel, T.: Weak-noise limit of Fokker-Planck models and non-differentiable potentials for dissipative dynamical systems. Phys. Rev. **31**, 1109 (1985)
237. Gregor, T., Tank, D.W., Wieschaus, E.F., Bialek, W.: Probing the limits to positional observation. Cell **130**, 153–164 (2007)
238. Greive, S.J., von Hippel, P.H.: Thinking quantitatively about transcriptional regulation. Nat. Rev. Mol. Cell Biol. **6**, 221–232 (2005)
239. Griffith, J.S.: Mathematics of cellular control processes. I. Negative feedback to one gene. J. Theor. Biol. **20**, 202–208 (1968)
240. Griffith, J.S.: Mathematics of cellular control processes. II. Positive feedback to one gene. J. Theor. Biol. **20**, 209–216 (1968)
241. Grill, S.W., Kruse, K., Julicher, F.: Theory of mitotic spindle oscillations. Phys. Rev. Lett. **94**, 108104 (2005)
242. Grimmett, G.R., Stirzaker, D.R.: Probability and Random Processes, 3rd edn. Oxford University Press, Oxford (2001)
243. Groff, J.R., DeRemigio, H., Smith, G.D.: Markov chain models of ion channels and calcium release sites, chap. 2. In: Stochastic Methods in Neuroscience, pp. 29–64. Oxford University Press, Oxford (2009)
244. Gross, S.P.: Hither and yon: a review of bi-directional microtubule-based transport. Phys. Biol. **1**, R1–R11 (2004)
245. Guerin, T., Prost, J., Martin, P., Joanny, J.-F.: Coordination and collective properties of molecular motors: theory. Curr. Opin. Cell Biol. **22**, 14–20 (2010)
246. Gundelfinger, E.D., Kessels, M.M., Qualmann, B.: Temporal and spatial coordination of exocytosis and endocytosis. Nat. Rev. Mol. Cell Biol. **4**, 127–139 (2003)
247. Gupta, A.: Stochastic model for cell polarity. Ann. Appl. Probab. **22**, 827–859 (2012)
248. Gupta, M.L., Carvalho, P., Roof, D.M., Pellman, D.: Plus end-specific depolymerase activity of Kip3, a kinesin-8 protein, explains its role in positioning the yeast mitotic spindle. Nat. Cell Biol. **8**, 913–923 (2006)
249. Halford, S.E.: An end to 40 years of mistakes in DNA-protein association kinetics? Biochem. Soc. Trans. **37**, 343–348 (2009)
250. Halford, S.E., Marko, J.F.: How do site-specific DNA-binding proteins find their targets? Nucleic Acid Res. **32**, 3040–3052 (2004)
251. Hall, A.: Rho GTPases and the actin cytoskeleton. Science **279**, 509–514 (1998).

252. Hanggi, P., Grabert, H., Talkner, P., Thomas, H.: Bistable systems: master equation versus Fokker–Planck modeling. Phys. Rev. A **29**, 371–378 (1984)
253. Hanggi, P., Talkner, P., Borkovec, M.: Reaction rate theory: fifty years after Kramers. Rev. Mod. Phys. **62**, 251–341 (1990)
254. Hawkins, R.J., Benichou, O., Piel, M., Voituriez, R.: Rebuilding cytoskeleton roads: active-transport-induced polarization of cell. Phys. Rev. E **80**, 040903(R) (2009)
255. He, G., Qian, H., Qian, M.: Stochastic theory of nonequilibrium steady states and its applications. Part II. Phys. Rep. **510**, 87–118 (2012)
256. Heinrich, R., Rapoport, T.A.: Generation of nonidentical compartments in vesicular transport systems. J. Cell Biol. **168**, 271–280 (2005)
257. Henley, J.M., Barker, E.A., Glebov, O.O.: Routes, destinations and delays: recent advances in AMPA receptor trafficking. Trends Neurosci. **34**, 258–268 (2011)
258. Herbert, K., La Porta, A., Wong, B., Mooney, R., Neuman, K., Landick, R., Block, S.: Sequence-resolved detection of pausing by single RNA polymerase molecules. Cell **125**, 1083–1094 (2006)
259. Higham, D.J.: An algorithmic introduction to numerical simulation of stochastic differential equations. SIAM Rev. **43**, 525–546 (2001)
260. Hill, T.L.: Theoretical problems related to the attachment of microtubules to kinetochores. Proc. Natl. Acad. Sci. USA **82**, 4404–4408 (1985)
261. Hille, B.: Ionic Channels of Excitable Membranes, 3rd edn. Sinauer Associates, Massachusetts (2001)
262. Hillen, T., Othmer, H.: The diffusion limit of transport equations derived from velocity-jump processes. SIAM J. Appl. Math. **61**, 751–775 (2000)
263. Hinch, R.: A mathematical analysis of the generation and termination of calcium sparks. Biophys. J. **86**, 1293–1307 (2004)
264. Hinch, R., Chapman, S.J.: Exponentially slow transitions on a Markov chain: the frequency of calcium sparks. Eur. J. Appl. Math. **16**, 427–446 (2005)
265. Hodgkin, A.L., Huxley, A.F.: A quantitative description of membrane and its application to conduction and excitation in nerve. J. Physiol. **117**, 500–544 (1952)
266. Holcman, D.: Modeling DNA and virus trafficking in the cell cytoplasm. J. Stat. Phys. **127**, 471–494 (2007)
267. Holcman, D., Schuss, Z.: Escape through a small opening: receptor trafficking in a synaptic membrane. J. Stat. Phys. **117**, 975–1014 (2004)
268. Holcman, D., Schuss, Z.: Control of flux by narrow passages and hidden targets in cellular biology. Rep. Prog. Phys. **76**, 074601 (2013)
269. Holcman, D., Triller, A.: Modeling synaptic dynamics driven by receptor lateral diffusion. Biophys. J. **91**, 2405–2415 (2006)
270. Holmes, W.R., Edelstein-Keshet, L.: A comparison of computational models for eukaryotic cell shape and motility. PLoS Comput. Biol. **8**, e1002793 (2012)
271. Holmes, W.R., Lin, B., Levchemko, A., Edelstein-Keshet, L.: Modelling cell polarization driven by synthetic spatially graded Rac activation. PLoS Comput. Biol. **8**, e1002366 (2012)
272. Holy, T.E., Leibler, S.: Dynamic instability of microtubules as an efficient way to search in space. Proc. Natl. Acad. Sci. USA **91**, 5682–5685 (1994)
273. Hopfield, J.J.: Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. Proc. Natl. Acad. Sci. USA **71**, 4135–4139 (1974)
274. Hough, L.E., Schwabe, A., Glaser, M.A., McIntosh, J.R. Betterton, M.D.: Microtubule depolymerization by the kinesin-8 motor Kip3p: a mathematical model. Biophys. J. **96**, 3050–3064 (2009)
275. Howard, J.: Mechanics of Motor Proteins and the Cytoskeleton. Sinauer, Sunderland (2001)
276. Howard, M.: How to build a robust intracellular concentration gradient. Trends Cell Biol. **22**, 311–317 (2012)
277. Howard, M., Kruse, K.: Cellular organization by self-organization: mechanisms and models for Min protein dynamics. J. Cell Biol. **168**, 533–536 (2005)

278. Howard, M., Rutenberg, A.D.: Pattern formation inside bacteria: fluctuations due to the low copy number of proteins. Phys. Rev. Lett. **90**, 128102 (2003)
279. Howard, M., Rutenberg, A.D., de Vet, S.: Dynamic compartmentalization of bacteria: accurate division in *E. coli*. Phys. Rev. Lett. **87**, 278102 (2001)
280. Howell, A.S., Savage, N.S., Johnson, S.A., Bose, I., Wagner, A.W., Zyla, T.R., Nijhout , H.F., Reed, M.C., Goryachev, A.B., Lew, D.J.: Singularity in polarization: rewiring yeast cells to make two buds. Cell **139**, 731–743 (2009)
281. Hu, T., Grossberg, A.Y., Shklovskii, B.I.: How proteins search for their specific sites on DNA: the role of DNA conformation. Biophys. J. **90**, 2731–2744 (2006)
282. Hu, J., Matzavinos, A., Othmer, H.G.: A theoretical approach to actin filament dynamics. J. Stat. Phys. **128**, 111–138 (2007)
283. Hu, J., Kang, H.-W., Othmer, H.G.: Stochastic analysis of reaction–diffusion processes. Bull. Math. Biol. **76**, 854–894 (2014)
284. Huang, K.C., Meir, Y, Wingreen, N.S.: Dynamic structures in Escherichia coli: spontaneous formation of MinE rings and MinD polar zones. Proc. Natl. Acad. Sci. USA **100**, 12724–12728 (2003)
285. Huber, F., Schnauss, J., Ronicke, S., Rauch, P., Muller, K., Futterer, C., Kas, J.E.: Emergent complexity of the cytoskeleton: from single filaments to tissue. Adv. Phys. **62**, 1–12 (2013)
286. Hudspeth, A.J.: Mechanical amplification of stimuli by hair cells. Curr. Opin. Neurobiol. **7**, 480–486 (1997)
287. Hudspeth, A.J.: Making an effort to listen: mechanical amplification in the ear. Neuron **59**, 530–545 (2008)
288. Huet, S., Karatekin, E., Tran, V.S., Fanget, I., Cribier, S., Henry, J.P.: Analysis of transient behavior in complex trajectories: application to secretory vesicle dynamics. Biophys. J. **91**, 3542–3559 (2006)
289. Hughes, B.D.: Random Walks and Random Environments Volume 1: Random Walks. Oxford University, Oxford (1995)
290. Huxley, A.F.: Muscle structure and theories of contraction. Prog. Biophys. Biophys. Chem. **7**, 255–318 (1957)
291. Inoue, S., Salmon, E.D.: Force generation by microtubule assembly/disassembly in mitosis and related movements. Mol. Cell Biol. **6** 1619–1640 (1995)
292. Isaacson, S.A.: The reaction-diffusion master equation as an asymptotic approximation of diffusion to a small target. SIAM J. Appl. Math. **7**, 77–111 (2009)
293. Isaacson, S., Peskin, C.: Incorporating diffusion in complex geometries into stochastic chemical kinetics simulations. SIAM J. Sci. Comput. **28**, 47–74 (2006)
294. Isaacson, S.A., McQueen, D.M., Peskin, C.S.: The influence of volume exclusion by chromatin on the time required to find specific DNA binding sites by diffusion. Proc. Natl. Acad. Sci. USA **108**, 3815–3820 (2011)
295. Jackson, M.B.: Molecular and Cellular Biophysics. Cambridge University Press, Cambridge (2006)
296. Jacob, F., Monod, J.: Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. **3**, 318–356 (1961)
297. Jacobs, M.H.: Diffusion Processes. Springer, New York (1967)
298. Jacobs, K.: Stochastic Processes for Physicists. Cambridge University Press, Cambridge (2010)
299. Jacobson, K., Mouritsen, O.G., Anderson, R.G.W.: Lipid rafts: at a crossroad between cell biology and physics. Nat. Cell Biol. **9**, 7–14 (2007)
300. Jilkine, A., Edelstein-Keshet, L.: A comparison of mathematical models for polarization of single eukaryotic cells in response to guided cues. PLoS Comput. Biol. **7**, e1001121 (2011)
301. Jilkine, A., Maree, A.F.M., Edelstein-Keshet, L.: Mathematical model for spatial segregation of the Rho-family GTPases based on inhibitory crosstalk. Bull. Math. Biol. **68**, 1169–1211 (2007)

302. Jilkine, A., Angenent, S.B., Wu, L.F., Altschuler, S.J.: A density-dependent switch drives stochastic clustering and polarization of signaling molecules. PLoS Comput. Biol. **7**, e1002271 (2011)
303. Joglekar, A.P., Hunt, A.J.: A Simple, mechanistic model for directional instability during mitotic chromosome movements. Biophys. J. **83**, 42–58 (2002)
304. Johann, D., Erlenkamper, C., Kruse, K.: Length regulation of active biopolymers by molecular motors. Phys. Rev. Lett. **108**, 258103 (2012)
305. Johnson, J.M., Jin, M., Lew, D.J.: Symmetry breaking and the establishment of cell polarity in budding yeast. Curr. Opin. Genet. Dev. **21**, 740–746 (2011)
306. Jose, J.V., Saletan, E.J.: Classical Dynamics: A Contemporary Approach. Cambridge University Press, Cambridge (2013)
307. Julicher, F., Prost, J.: Cooperative molecular motors. Phys. Rev. Lett. **75**, 2618–2621 (1995)
308. Julicher, F., Prost, J.: Spontaneous oscillations of collective molecular motors. Phys. Rev. Lett. **78**, 4510–4513 (1997)
309. Julicher, F., Ajdari, A., Prost, J.: Modeling molecular motors. Rev. Mod. Phys. **69**, 1269–1281 (1997)
310. Jung, P., Brown, A.: Modeling the slowing of neurofilament transport along the mouse sciatic nerve. Phys. Biol. **6**, 046002 (2009)
311. Kac, M.: A stochastic model related to the telegrapher's equation. Rocky Mountain J. Math. **3**, 497–509 (1974)
312. Kaern, M., Elston, T.C., Blake, W.J., Collins, J.J.: Stochasticity in gene expression: from theories to phenotypes. Nat. Rev. Genet. **6**, 451–464 (2005)
313. Kahana, A., Kenan, G., Feingold, M., Elbaum, M., Granek, R.: Active transport on disordered microtubule networks: the generalized random velocity model. Phys. Rev. E **78**, 051912 (2008)
314. Kaizu, K., de Ronde, W., Paijmans, J., Tkahashi, K., Tostevin, F., ten Wolde, P. R.: The Berg-Purcell limit revisited. Biophys. J. **106**, 976–985 (2014)
315. Kalab, P., Weis, K., Heald, R:. Visualization of a Ran-GTP gradient in interphase and mitotic Xenopus egg extracts. Science **295**, 2452–2456 (2002)
316. Kalinay, P., Percus, J.K.: Corrections to the Fick-Jacobs equation. Phys. Rev. E **74**, 041203 (2006)
317. Kardar, M.: Statistical Physics of Particles. Cambridge University Press, Cambridge (2007)
318. Karmakar, R., Bose, I.: Graded and binary responses in stochastic gene expression. Phys. Biol. **1**, 197–204 (2004)
319. Katsura, I.: Determination of bacteriophage λ tail length by a protein ruler. Nature **327**, 73–75 (1987)
320. Keener, J.P.: How Salmonella typhimurium measures the length of flagellar filaments. Bull. Math. Biol. **68**, 1761–1778 (2006)
321. Keener, J.P., Newby, J.M.: Perturbation analysis of spontaneous action potential initiation by stochastic ion channels. Phy. Rev. E **84**, 011918 (2011)
322. Keener, J.P., Sneyd, J.: Mathematical Physiology I: Cellular Physiology, 2nd edn. Springer, New York (2009)
323. Keizer, J.: Nonequilibrium statistical thermodynamics and the effect of diffusion on chemical reaction rates. J. Phys. Chem. **86**, 5052–5067 (1982)
324. Keizer, J., Smith, G.D.: Spark-to-wave transition: saltatory transmission of calcium waves in cardiac myocytes. Biophys. Chem. **72**, 87–100 (1998)
325. Keller, D., Bustamante, C.: The mechanochemistry of molecular motors. Biophys. J. **78**, 541–556 (2000)
326. Keller, E., Segel, L.: Initiation of slime mold aggregation viewed as an instability. J. Theor. Biol. **26**, 399–415 (1970)
327. Kenkre, V.M., Giuggioli, L., Kalay, Z.: Molecular motion in cell membranes: analytical study of fence-hindered random walks. Phys. Rev. E **77**, 051907 (2008)
328. Kepler, T.B., Elston, T.C.: Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. Biophys. J. **81**, 3116–3136 (2001)

329. Keren, K., Pincus, Z., Allen, G.M., Barnhart, E.L., Marriott, G., Mogilner, A., Theriot, J.A.: Mechanism of shape determination in motile cells. Nature **453**, 475–481 (2008)
330. Kerr, J.M., Blanpied, T.A.: Subsynaptic AMPA receptor distribution is acutely regulated by actin-driven reorganization of the postsynaptic density. J. Neurosci. **32**, 658–673 (2012)
331. Kholodenko, B.N.: Spatially distributed cell signalling. FEBS Lett. **583**, 4006–4012 (2009)
332. Kifer, Y.: Large deviations and adiabatic transitions for dynamical systems and Markov processes in fully coupled averaging. Memoirs of the AMS **201**, issue 944 (2009)
333. Kimmel, M., Axelrod, D.E.: Mathematical models of gene amplification with applications to cellular drug resistance and tumorigenicity. Genetics **125**, 633–644 (1990)
334. Kimmel, M., Axelrod, D.E.: Branching Processes in Biology. Springer, New York (2002)
335. King, D.P., Takahashi, J.S.: Molecular genetics of circadian rhythms in mammals. Ann. Rev. Neurosci. **23**, 713–742 (2000)
336. Kirschner, M., Mitchison, T.: Beyond self-assembly: from microtubules to morphogenesis. Cell **45**, 329–342 (1986)
337. Klann, M., Koeppl, H., Reuss, M.: Spatial modeling of vesicle transport and the cytoskeleton: the challenge of hitting the right road. PLoS One **7**, e29645 (2012)
338. Kloeden, P.E., Platen, E.: Numerical Solution of Stochastic Differential Equations. Springer, New York (2000)
339. Klumpp, S., Lipowsky, R.: Traffic of molecular motors through tube-like compartments. J. Stat. Phys. **113**, 233–268 (2003)
340. Klumpp, S., Lipowsky, R.: Cooperative cargo transport by several molecular motors. Proc. Natl. Acad. Sci. USA **102**, 17284–17289 (2005)
341. Knessl, C., Matkowsky, B.J., Schuss, Z., Tier, C.: An asymptotic theory of large deviations for Markov jump processes. SIAM J. Appl. Math. **46**, 1006–1028 (1985)
342. Knowles, R.B., Sabry, J.H., Martone, M.E., Deerinck, T.J., Ellisman, M.H., Bassell, G.J., Kosik, K.S.: Translocation of RNA granules in living neurons. J. Neurosci. **16**, 7812–7820 (1996)
343. Koeppl, H., Hafner, M., Ganguly, A., Mehrotra, A.: Deterministic characterization of phase noise in biomolecular oscillators. Phys. Biol. **8**, 055008 (2011)
344. Kolomeisky, A.B.: Asymmetric simple exclusion model with local inhomogeneity. J. Phys. A **31**, 1153–1164 (1998)
345. Kolomeisky, A.B.: Channel-facilitated molecular transport across membranes: attraction, repulsion and asymmetry. Phys. Rev. Lett. **98**, 048105 (2007)
346. Kolomeisky, A.B.: Physics of protein-DNA interactions: mechanisms of facilitated target search. Phys. Chem. Chem. Phys. **13**, 2088–2095 (2011)
347. Kolomeisky, A.B., Fisher, M.E.: A simple kinetic model describes the processivity of myosin-V. Biophys. J. **84**, 1642–1650 (2003)
348. Kolomeisky, A., Fisher, M.: Molecular motors: a theorist's perspective. Annu. Rev. Phys. Chem. **58**, 675–695 (2007)
349. Kolomeisky, A.B., Schutz, G.M., Kolomeisky, E.B., Straley, J.P.: Phase diagram of one-dimensional driven lattice gases with open boundaries. J. Phys. A **31**, 6911–6919 (1998)
350. Kondo S, Asai, R.: A reaction-diffusion wave on the skin of the marine angelfish Pomacanthus. Nature **376**, 765–768 (2002)
351. Kosik, K.S., Joachim, C.L., Selkoe, D.J.: Microtubule-associated protein tau is a major antigenic component of paired helical filaments in Alzheimer disease. Proc. Natl. Acad. Sci. USA **83**, 4044–4048 (1986)
352. Koster, G., VanDuijn, M., Hofs, B., Dogterom, M.: Membrane tube formation from giant vesicles by dynamic association of motor proteins. Proc. Natl. Acad. Sci. USA **100**, 15583–15588 (2003)
353. Krapivsky, P.L., Mallick, K.: Fluctuations in polymer translocation. J. Stat. Mech. **2010**, P07007 (2010)
354. Krapivsky, P.L., Redner, S., Ben-Naim, E.: A Kinetic View of Statistical Physics. Cambridge University Press, Cambridge (2010)

355. Krug, J.: Boundary-induced phase transitions in driven diffusive systems. Phys. Rev. E **67**, 1882–1185 (1991)
356. Kruse, K.: A dynamic model for determining the middle of *Escherichia coli*. Biophys. J. **82**, 618–627 (2002)
357. Kuan, H.-S., Betterton, M.D.: Biophysics of filament length regulation by molecular motors. Phys. Biol. **10**, 036004 (2013)
358. Kubo, R.: Fluctuation, relaxation and resonance in magnetic systems. In: TerHaar, D. (ed.) Stochastic Theory of Line Shape. Oliver and Boyd, Edinburgh (1962)
359. Kural, C., Kim, H., Syed, S., Goshima, G., Gelfand, V.I., Selvin, P.R.: Kinesin and dynein move a peroxisome in vivo: a tug-of-war or coordinated movement? Science **308**, 1469–1472 (2005)
360. Kurtz, T.G.: Limit theorems and diffusion approximations for density dependent Markov chains. Math. Prog. Stud. **5**, 67–78 (1976)
361. Kurtz, T.G.: Representations of Markov processes as multiparameter changes. Ann. Probab. **8**, 682–715 (1980)
362. Kustanovich, T., Rabin, Y.: Metastable network model of protein transport through nuclear pores. Biophys. J. **86**, 2008–2016 (2004)
363. Kusumi, A., Sako, Y., Yamamoto, M.: Confined lateral diffusion of membrane receptors as studied by single particle tracking (nanovid microscopy): effects of calcium-induced differentiation in cultured epithelial cells. Biophys. J. **65**, 2021–2040 (1993)
364. Kusumi, A., Nakada, C., Ritchie, K., Murase, K., et al.: Paradigm shift of the plasma membrane concept from the two-dimensional continuum fluid to the partitioned fluid: high-speed single-molecule tracking of membrane molecules. Annu. Rev. Biophys. Biomol. Struct. **34**, 351–354 (2005)
365. Kusumi, A., Shirai, Y.M., Koyama-Honda, I., Suzuki, K.G.N., Fujiwara, T.K.: Hierarchical organization of the plasma membrane: investigations by single-molecule tracking vs. fluorescence correlation spectroscopy. FEBS Lett. **584**, 1814–1823 (2010)
366. Kuznetzov, Y.A.: Elements of Applied Bifurcation Theory, 3rd edn. Springer, New York (2010)
367. Kuznetsov, A.V., Avramenko, A.A.: The method of separation of variables for solving equations describing molecular-motor-assisted transport of intracellular particles in a dendrite or axon. Proc. Roy. Soc. A **464**, 2867–2886 (2008)
368. Lagache, T., Holcman, D.: Effective motion of a virus trafficking inside a biological cell. SIAM J. Appl. Math. **68**, 1146–1167 (2009)
369. Lagache, T., Dauty, E., Holcman, D.: Physical principles and models describing intracellular virus particle dynamics. Curr. Opin. Microbiol. **12**, 439–445 (2009)
370. Lander, A.D.: Pattern, growth and control. Cell **144**, 955–969 (2011)
371. Lauffenburger, D.A.: Receptors: Models for Binding, Trafficking, and Signaling. Oxford University Press, Oxford (1996)
372. Lauffenburger, D.A., Horwitz, A.F.: Cell migration: a physically-integrated molecular process. Cell **84**, 359–369 (1996)
373. Lawson, M.J., Drawert, B., Khammash, M., Petzold, L., Yi, T-M.: Spatial stochastic dynamics enable robust cell polarization. PLoS Comput. Biol. **9**, e1003139 (2012)
374. Layton, A.T., Savage, N.S., Howell, A.S., Carroll, S.Y., Drubin, D.G., Lew, D.J.: Modeling vesicle traffic reveals unexpected consequences for cdc42p-mediated polarity establishment. Curr. Biol. **21**, 184–194 (2011)
375. Lebowitz, J.L., Percus, J.K.: Kinetic equations and density expansions: exactly solvable one-dimensional system. Phys. Rev. **155**, 122–138 (1967)
376. Lechleiter, J., Girard, S., Peralta, E., Clapham, D.: Spiral calcium wave propagation and annihilation in *Xenopus laevis* oocytes. Science **252**, 123–126 (1991)
377. Leduc, C., et al.: Cooperative extraction of membrane nanotubes by molecular motors. Proc. Natl. Acad. Sci. USA **101**, 17096–17101 (2004)
378. Lefebvre, P.A., Rosenbaum, J.L.: Regulation of the synthesis and assembly of ciliary and flagellar proteins during regeneration. Annu. Rev. Cell Biol. **2**, 517–546 (1986)

379. Leibler, S., Huse, D.A.: Porters versus rowers: a unified stochastic model of motor proteins. J. Cell Biol. **121**, 1357–1368 (1993)
380. Leitner, D.M. , Brown, F.L.H., Wilson, K.R.: Regulation of protein mobility in cell membranes: a dynamic corral model. Biophys. J. **78**, 125–135 (2000)
381. Leloup, J.-C., Goldbeter, A.: Toward a detailed computational model for the mammalian circadian clock. Proc. Natl. Acad. Sci. USA **100**, 7051–7056 (2003)
382. Leloup, J.-C., Gonze, D., Goldbeter, A.: Limit cycle models for circadian rhythms based on transcriptional regulation in Neurospora and Drosophila. J. Biol. Rhythms **14**, 433–448 (1999)
383. Lestas, I., Vinnicombe, G., Paulsson, J.: Fundamental limits on the suppression of molecular fluctuations. Nature **467**, 174–178 (2010)
384. Levchenko, A., Iglesias, P.A.: Models of eukaryotic gradient sensing: application to chemotaxis of amoebae and neutrophils. Biophys. J. **82**, 50–63 (2002)
385. Levin, S.A.: The problem of pattern and scale in ecology. Ecology **73**, 1943–1967 (1992)
386. Levitt, D.G.: Dynamics of a single-file pore: non-Fickian behavior. Phys. Rev. A **8**, 3050–3054 (1973)
387. Levitt, D.G.: Modeling of ion channels. J. Genet. Physiol. **113**, 789–794 (1999)
388. Li, Y., Rinzel, J.: Equations for $InsP_3$ receptor-mediated calcium oscillations derived from a detailed kinetic model: a Hodgkin-Huxley like formalism. J. Theor. Biol. **166**, 461–473 (1994)
389. Li, G.W., Berg, O.G., Elf, J.: Effects of macromolecular crowding and DNA looping on gene regulation kinetics. Nat. Phys. **4**, 294–297 (2009)
390. Liepelt, S., Lipowsky, R.: Kinesin's network of chemomechanical motor cycles. Phys. Rev. Lett. **98**, 258102 (2007)
391. Lipowsky, R., Klumpp, S.: 'Life is motion': multiscale motility of molecular motors. Phys. A Stat. Mech. Appl. **352**, 53–112 (2005)
392. Lippincott-Schwartz, J., Phair, R.D.: Lipids and cholesterol as regulators of traffic in the endomembrane system. Annu. Rev. Biophys. **39**, 559–578 (2010)
393. Lippincott-Schwartz, J., Roberts, T.H., Hirschberg, K.: Secretory protein trafficking abd organelle dynamics in living cells. Annu. Rev. Cell Dev. Biol. **16**, 557–589 (2000)
394. Liu, J., Deasi, A., Onunchic, J.N., Hwa, T.: An integrated mechanobiochemical feedback mechanism describes chromosome motility from prometaphase to anaphase in mitosis. Proc. Natl. Acad. Sci. USA **105**, 13752–13757 (2008)
395. Logan, J.D.: Applied Mathematics, 4th edn. Wiley, New Jersey (2013)
396. Loose, M., Kruse, K., Schwille, P.: Protein self-organization: lessons from the Min system. Annu. Rev. Biophys. **40**, 315–336 (2011)
397. Loverdo, C., Benichou, O., Moreau, M., Voituriez, R.: Enhanced reaction kinetics in biological cells. Nat. Phys. **4**, 134–137 (2008)
398. Loverdo, C., Benichou, O., Moreau, M., Voituriez, R.: Robustness of optimal intermittent search strategies in one, two, and three dimensions. Phys. Rev. E **80**, 031146 (2009)
399. Lugo, C.A., McKane, A.J.: Quasi-cycles in a spatial predator-prey model. Phys. Rev. E **78**, 051911 (2008)
400. Luo, L.: Actin cytoskeleton regulation in neuronal morphogenesis and structural plasticity. Annu. Rev. Neurosci. **18**, 601–635 (2002)
401. Luo, C.H., Rudy, Y.: A dynamic model of the cardiac ventricular action potential. II. Afterdepolarizations, triggered activity, and potentiation. Circ. Res. **74**, 1097–1113 (1994)
402. Lyons, L.: Bayes and Frequentism: a particle physicist's perspective. Contemp. Phys. **54**, 1–16 (2013)
403. Macara, I.G.: Transport into and out of the nucleus. Microbiol. Mol. Biol. Rev. **65**, 570–594 (2001)
404. MacDonald, C.T., Gibbs, J.H.: Concerning the kinetics of polypeptide synthesis on polyribosomes. Biopolymers **7**, 707–725 (1969)
405. MacDonald, C.T., Gibbs, J.H., Pipkin, A.C.: Kinetics of biopolymerization on nucleic acid templates. Biopolymers **6**, 1–25 (1968)

406. MacGillavry, H.D., Kerr, J.M., Blanpied, T.A.: Lateral organization of the postsynaptic density. Mol. Cell. Neurosci. **48**, 321–331 (2011)
407. Mackey, M.C., Tyran-Kaminska, M., Yvinec, R.: Dynamic behavior of stochastic gene expression in the presence of bursting. SIAM J. Appl. Math. **73**, 1830–1852 (2013)
408. Maheshri, N., O'Shea, E.K.: Living with noisy genes: how cells function reliably with inherent variability in gene expression. Annu. Rev. Biophys. Biomol. Struct. **36**, 413–434 (2007)
409. Maier, R.S., Stein, D.L.: Limiting exit location distribution in the stochastic exit problem. SIAM J. Appl. Math. **57**, 752–790 (1997)
410. Maly, I.V., Borisy, G.G.: Self-organization of a propulsive actin network as an evolutionary process. Proc. Natl. Acad. Sci. USA **98**, 11324–11329 (2001)
411. Mandelbrot, B.B., Van Ness, J.W.: Fractional Brownian motions, fractional noise and applications. SIAM Rev. **10**, 422–437 (1968)
412. Manley, G.A.: Evidence for an active process and a cochlear amplifier in nonmammals. J. Neurophysiol. **86**, 541–549 (2001)
413. Maoileidigh, D.O., Tadigotla, V.R., Nudler, E., Ruckenstein, A. E.: A unified model of transcription elongation: what have we learned from single-molecule experiments? Biophys. J. **100**, 1157–1166 (2011)
414. Marban, E., Robinson, S.W., Wier, W.G. : Mechanisms of arrhythmogenic delayed and early afterdepolarizations in ferret ventricular muscle. J. Clin. Invest. **78**, 1185–1192 (1986)
415. Marco, E., Wedlich-Soldner, R., Li, R., Altschuler, S.J., Wu, L.F.: Endocytosis optimizes the dynamic localization of membrane proteins that regulate cortical polarity. Cell **129**, 411–422 (2007)
416. Marcy, Y., Prost, J., Carlier, M.F., Sykes, C.: Forces generated during actin-based propulsion: a direct measurement by micromanipulation. Proc. Natl. Acad. Sci. USA **20**, 5992–5997 (2004)
417. Maree, A.F., Jilkine, A., Dawes, A., Grieneisen, V.A., Edelstein-Keshet, L.: Polarization and movement of keratocytes: a multiscale modelling approach. Bull. Math. Biol. **68**, 1169–1211 (2006)
418. Markin, V.S., Hudspeth, A.J.: Gating-spring models of mechanoelectrical transduction by hair cells of the internal ear. Annu. Rev. Biophys. Biomol. Struct. **24**, 59–83 (1995)
419. Marshall, W.F., Rosenbaum, J.L.: Intraflagellar transport balances continuous turnover of outer doublet microtubules: implications for flagellar length control. J. Cell Biol. **155**, 405–414 (2001)
420. Marshall, W.F., Qin, H., Rodrigo Brenni, M., Rosenbaum, J.L.: Flagellar length control system: testing a simple model based on intraflagellar transport and turnover. Mol. Biol. Cell **16**, 270–278 (2005)
421. Martin, S.G., Berthelot-Grosjean, M.: Polar gradients of the DYRK-family kinase Pom1 couple cell length with the cell cycle. Nature **459**, 852–856 (2009)
422. Martin, P., Mehta, A.D., Hudspeth, A.J.: Negative hair-bundle stiffness betrays a mechanism for mechanical amplification by the hair cell. Proc. Natl. Acad. Sci. USA **97**, 12026–12031 (2000)
423. Martin, P., Hudspeth, A.J., Julicher, F.: Comparison of a hair bundle's spontaneous oscillations with its response to mechanical stimulation reveals the underlying active process. Proc. Natl. Acad. Sci. USA **98**, 14380–14385 (2001)
424. Martin, P., Bozovic, D., Choe, Y., Hudspeth, A.J.: Spontaneous oscillation by hair bundles of the bullfrog's sacculus. J. Neurosci. **23**, 4533–4454 (2003)
425. Marzen, S., Garcia, H.G., Phililps, R.: Statistical mechanics of Monod-Wyman-Changeux (MWC) models. J. Mol. Biol. **425**, 1433–1460 (2013)
426. Matkowsky, B.J., Schuss, Z.: The exit problem for randomly perturbed dynamical systems. SIAM J. Appl. Math. **33**, 365–382 (1977)
427. McAdams, H.H., Arkin, A.: Stochastic mechanisms in gene expression. Proc. Natl. Acad. Sci. USA **94**, 814–819 (1997)
428. McIntosh, J.R., Molodtsov, M.I., Ataullakhanov, F.I.: Biophysics of mitosis. Q. Rev. Biophys. **45**, 147–207 (2012)

429. McKane, A.J., Biancalani, T., Rogers, T.: Stochastic pattern formation and spontaneous polarization: the linear noise approximation and beyond. Bull. Math. Biol. **76**, 895–921 (2014)
430. McKeithan, K.: Kinetic proofreading in T-cell receptor signal transduction. Proc. Natl. Acad. Sci. USA **92**, 5042–5046 (1995)
431. Mclean, D.R., Graham, B.P.: Mathematical formulation and analysis of a continuum model for tubulin-driven neurite elongation. Proc. Roy. Soc. Lond. A **460**, 2437–2456 (2004)
432. Meacci, G., Kruse, K.: Min-oscillations in *Escherichia coli* induced by interactions of membrane-bound proteins. Phys. Biol. **2**, 89–97 (2005)
433. Mehta, A.D., Rock, R.S., Rief, M., Spudich, J.A., Mooseker, M.S., Cheney, R.E.: Myosin-V is a processive actin-based motor. Nature **400**, 590–593 (2009)
434. Meinhardt, M., de Boer, P.A J.: Pattern formation in *Escherichia coli*: a model for the pole-to-pole oscillations of min proteins and the localization of the division site. Proc. Natl. Acad. Sci. USA **98**, 14202–14207 (2001)
435. Melbinger, A., Reese, L., Frey, E.: Microtubule length regulation by molecular motors. Phys. Rev. Lett. **108**, 258104 (2012)
436. Mello, B.A., Tu, Y.: Quantitative modeling of sensitivity in bacterial chemotaxis: the role of coupling among different chemoreceptor species. Proc. Natl. Acad. Sci. USA **100**, 8223–8228 (2003)
437. Menchon, S.A., Gartner, A., Roman, P., Dotti, C.G.: Neuronal (bipolarity) as a self-organized process enhanced by growing membrane. PLoS One **6**, e24190 (2011)
438. Metzler, R., Klafter, J.: The random walk's guide to anomalous diffusion: a fractional dynamics approach. Phys. Rep. **339**, 1–77 (2000)
439. Meyers, J., Craig, J., Odde, D.J.: Potential for control of signaling pathways via cell size and shape. Curr. Biol. **16**, 1685–1693 (2006).
440. Mirny, L., Slutsky, M., Wunderlich, Z., Tafvizi, A., Leith, J., Kosmrlj, A.: How a protein searches for its site on DNA: the mechanism of facilitated diffusion. J. Phys. A **42**, 434013 (2009)
441. Misteli, T.: Self-organization in cell architecture. J. Cell Biol. **155**, 181–186 (2001)
442. Mitchison, T.J.: Self-organization of polymer-motor systems in the cytoskeleton. Phil. Trans. Roy. Soc. Lond. B Biol. Sci. **336**, 99–106 (1992)
443. Mitchison, T.J., Kirschner, M.W.: Dynamic instability of microtubule growth. Nature **312**, 237–242 (1984)
444. Mogilner, A.: Mathematics of cell motility: have we got its number? J. Math. Biol. **58**, 105–134 (2009)
445. Mogilner, A., Craig, E.: Towards a quantitative understanding of mitotic spindle assembly and mechanics. J. Cell Sci. **123**, 3435–3445 (2010)
446. Mogilner, A., Edelstein-Keshet, L.: Regulation of actin dynamics in rapidly moving cells: a quantitative analysis. Biophys. J. **83**, 1237–1258 (2002)
447. Mogilner, A., Oster, G.: Cell motility driven by actin polymerization. Biophys. J. **71**, 3030–3045 (1996)
448. Mogilner, A., Oster, G.: Force generation by actin polymerization II: the elastic ratchet and tethered filaments. Biophys. J. **84**, 1591–1605 (2003)
449. Mogilner, A., Fisher, A.J., Baskin, R.: Structural changes in the neck linker of kinesin explain the load dependence of the motor's mechanical cycle. J. Theor. Biol. **211**, 143–157 (2001)
450. Mogilner, A., Wollman, R., Civelekoglu-Sholey, G., Scholey, J.: Trends Cell Biol. **16**, 89–96 (2006)
451. Monod, J., Wyman, J., Changeux, J.-P.: On the nature of allosteric transitions: a plausible model. J. Mol. Biol. **12**, 88–118 (1965)
452. Mori, Y., Jilkine, A., Edelstein-Keshet, L.: Wave-pinning and cell polarity from a bistable reaction-diffusion system. Biophys. J. **94**, 3684–3697 (2008)
453. Mori, Y., Jilkine, A., Edelstein-Keshet, L.: Asymptotic and bifurcation analysis of wave-pinning in a reaction-diffusion model for cell polarization. SIAM J. Appl. Math. **71**, 1401–1427 (2011)

454. Morris, C., Lecar, H.: Voltage oscillations in the barnacle giant muscle fiber. J. Biophys. **35**, 193–213 (1981)
455. Moseley, J.B., Mayeux, A., Paoletti, A., Nurse, P.: A spatial gradient coordinates cell size and mitotic entry in fission yeast. Nature **459**, 857–860 (2009)
456. Muller, M.J.I., Klumpp, S., Lipowsky, R.: Motility states of molecular motors engaged in a stochastic tug-of-war. J. Stat. Phys. **133**, 1059–1081 (2008)
457. Muller, M.J.I., Klumpp, S., Lipowsky, R.: Tug-of-war as a cooperative mechanism for bidirectional cargo transport by molecular motors. Proc. Natl. Acad. Sci. USA **105**, 4609–4614 (2008)
458. Mullins, R.D., Heuser, J.A., Pollard, T.D.: The interaction of Arp2/3 complex with actin: nucleation, high affinity pointed end capping, and formation of branching networks of filaments. Proc. Natl. Acad. Sci. USA **95**, 6181–6186 (1998)
459. Munoz-Garcia, J., Kholodenko, B.N.: Signaling and control from a systems perspective. Biochem. Soc. Trans. **38**, 1235–1241 (2010)
460. Munoz-Garcia, J., Neufeld, Z., Kholodenko, B.N.: Positional information generated by spatially distributed signaling cascades. PLoS Comput. Biol. **3**, e1000330 (2009)
461. Murray, J.D.: Mathematical Biology, vols. I and II, 3rd edn. Springer, Berlin (2008)
462. Muthukumar, M.: Polymer translocation through a hole. J. Chem. Phys. **111**, 10371–10374 (1999)
463. Muthukumar, M.: Polymer Translocation. CRC Press, Boca Raton (2011)
464. Nadler, B., Schuss, Z., Singer, A., Eisenberg, R.S.: Ionic diffusion through confined geometries: from Langevin equations to partial differential equations. J. Phys. Condens. Matter **16**, S2153–S2165 (2004)
465. Nadrowski, B., Martin, P., Julicher, F.: Active hair-bundle motility harnesses noise to operate near an optimum of mechanosensitivity. Proc. Natl. Acad. Sci. USA **101**, 12195–12200 (2004)
466. Naeh, T., Klosek, M.M., Matkowsky, B.J., Schuss, Z.: A direct approach to the exit problem. SIAM J. Appl. Math. **50**, 595–627 (1990)
467. Nedelec, F., Surrey, T., Maggs, A.C., Leibler, S.: Self-organization of microtubules and motors. Nature **389**, 305–308 (1997)
468. Neukirchen, D., Bradke, F.: Neuronal polarization and the cytoskeleton. Sem. Cell Dev. Biol. **22** .825–833 (2011)
469. Neuman, K.C., Abbondanzieri, E.A., Landick, R., Gelles, J., Block, S.M.: Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. Cell. **115**, 437–447 (2003)
470. Newby, J.M.: Isolating intrinsic noise sources in a stochastic genetic switch. Phys. Biol. **9**, 026002 (2012)
471. Newby, J.M., Bressloff, P.C.: Directed intermittent search for a hidden target on a dendritic tree. Phys. Rev. E **80**, 021913 (2009)
472. Newby, J.M., Bressloff, P.C.: Local synaptic signalling enhances the stochastic transport of motor-driven cargo in neurons. Phys. Biol. **7**, 036004 (2010)
473. Newby, J.M., Bressloff, P.C.: Quasi-steady state reduction of molecular-based models of directed intermittent search. Bull. Math. Biol. **72**, 1840–1866 (2010)
474. Newby, J.M., Bressloff, P.C.: Random intermittent search and the tug-of-war model of motor-driven transport. J. Stat. Mech. **2010**, P04014 (2010)
475. Newby, J.M., Keener, J.P.: An asymptotic analysis of the spatially inhomogeneous velocity-jump process. SIAM Multiscale Model. Simul. **9**, 735–765 (2011)
476. Newby, J.M., Bressloff, P.C., Keeener, J.P.: The effect of potassium channels on spontaneous action potential initiation by stochastic ion channels. Phys. Rev. Lett. **111**, 128101 (2013)
477. Nicolis, G., Prigogine, I.: Self-organization in nonequilibrium systems, 512 pp. Wiley, New York (1977)
478. Niethammer, P., Bastiaens, P., Karsenti, E.: Stathmin-tubulin interaction gradients in motile and mitotic cells. Science **303**, 1862–1866 (2004)
479. Ninio, J.: Kinetic amplification of enzyme discrimination. Biochimie **57**, 587–595 (1975)

480. Novak, I.L., Kraikivski, P., Slepchenko, B.M.: Diffusion in cytoplasm: effects of excluded volume due to internal membranes and cytoskeletal structures. Biophys. J. **97**, 758–767 (2009)
481. Nowak, S., Fok, P.W., Chou, T.: Dynamic boundaries in asymmetric exclusion processes. Phys. Rev. E **76**, 031135 (2007)
482. Nudler, E.: RNA polymerase backtracking in gene regulation and genome instability. Cell **149**, 1438–1445 (2012)
483. Oksendal, B.: Stochastic Differential Equations: An Introduction with Applications, 5th edn. Springer, New York (1998)
484. Onsum, M., Rao, C.V.: A mathematical model for neutrophil gradient sensing and polarization. PLoS Comput. Biol. **3**, e36 (2007)
485. Oshanin, O., Lindenberg, K., Wio, H.S., Burlatsky, S.: Efficient search by optimized intermittent random walks. J. Phys. A **42**, 434008 (2009)
486. Othmer, H.G., Hillen, T.: The diffusion limit of transport equations II: chemotaxis equations. SIAM J. Appl. Math. **62**, 1222–1250 (2002)
487. Othmer, H., Dunbar, S., Alt, W.: Models of dispersal in biological systems. J. Math. Biol. **26**, 263–298 (1988)
488. Othmer, H.G., Painter, K., Umulis, D., Xue, C.: The intersection of theory and application in elucidating pattern formation in developmental biology. Math. Model. Nat. Phenom. **4**, 1–80 (2009)
489. Otsuji, M., Ishihara, S., Co, C., Kaibuchi, K., Mochizuki, A., Kuroda, S.: A mass conserved reaction-diffusion system captures properties of cell polarity. PLoS Comput. Biol. **3**, e108 (2007)
490. Padte, N.N., Martin, S.G., Howard, M., Chang, F.: The cell-end factor pom1p inhibits mid1p in specification of the cell division plane in fission yeast. Curr. Biol. **16**, 2480–2487 (2006)
491. Pakdaman, K., Thieullen, M., Wainrib, G.: Fluid limit theorems for stochastic hybrid systems with application to neuron models. Adv. Appl. Probab. **42**, 761–794 (2010)
492. Panda, S., Hogenesch, J.B., Kay, S.A.: Circadian rhythms from flies to human. Nature **417**, 329–335 (2002)
493. Panja, D., Barkema, G.T., Kolomeisky, A.B.: Through the eye of the needle: recent advances in understanding biopolymer translocation. J. Phys. Condens. Matter **25**, 413101 (2013)
494. Papanicolaou, G.C.: Asymptotic analysis of transport processes. Bull. Am. Math. Soc. **81**, 330–392 (1975)
495. Parmeggiani, A., Julicher, F., Ajdari, A., Prost, J.: Energy transduction of isothermal ratchets: generic aspects and specific examples close to and far from equilibrium. Phys. Rev. E **60**, 2127–2140 (1999)
496. Parmeggiani, A., Franosch, T., Frey, E.: Phase coexistence in driven one-dimensional transport. Phys. Rev. Lett. **90**, 086601 (2003)
497. Parmeggiani, A., Franosch, T., Frey, E.: Totally asymmetric simple exclusion process with langmuir kinetics. Phys. Rev. E **70**, 046101 (2004)
498. Parker, I., Ivorra, I.: Localized all-or-none calcium liberation by inositol triphosphate. Science **250**, 977–979 (1990)
499. Parker, I., Yao, Y.: Regenerative release of calcium from functionally discrete subcellular stores by inositol triphosphate. Proc. Roy. Soc. Lond. B **246**, 269–274 (1991)
500. Parsons, J.T., Horwitz, A.R., Schwartz, M.A.: Cell adhesion: integrating cytoskeletal dynamics and cellular tension. Nat. Rev. Mol. Cell Biol. **11**, 633–643 (2010)
501. Partch, C.L., Green, C.B., Takahashi, J.S.: Molecular architecture of the mammalian circadian clock. Trends Cell Biol. **24**, 90–99 (2013)
502. Paulsson, J.: Models of stochastic gene expression. Phys. Life Rev. **2**, 157–175 (2005)
503. Peliti, L.: Path integral approach to birth-death processes on a lattice. J. Phys. **46**, 1469–1483 (1985)
504. Percus, J.K.: Anomalous self-diffusion for one-dimensional hard cores. Phys. Rev. A **9**, 557–559 (1974)

505. Perelson, A.P., Coutsias, E.A.: A moving boundary model of acrosomal elongation. J. Math. Biol. **23**, 361–379 (1986)
506. Perthame, B., Souganidis, P.E.: Asymmetric potentials and motor effect: a homogenization approach. Ann. I. H. Poincaré-AN **26**, 2055–2071 (2009)
507. Peskin, C.P., Odell, G.M., Oster, G.F.: Cellular motions and thermal fluctuations: the brownian ratchet. Biophys. J. **65**, 316–324 (1993)
508. Peskin, C.S., Ermentrout, B., Oster, G.: The correlation ratchet: a novel mechanism for generating directed motion by ATP hydrolysis. In: Mow, V.C., et al. (ed.) Cell Mechanics and Cellular Engineering. Springer, New-York (1995)
509. Phillips, R., Kondev, J., Theriot, J., Garcia, H.: Physical Biology of the Cell, 2nd edn. Garland Science, New York (2012)
510. Pierce-Shimomura, J., Morse, T., Lockery, S.: The fundamental role of pirouettes in C. elegans chemotaxis. J. Neurosci. **19**, 9557–9569 (1999)
511. Pillay, S., Ward, M.J., Peirce, A., Kolokolnikov, T.: An asymptotic analysis of the mean first passage time for narrow escape problems: Part I: two-dimensional domains. SIAM Multiscale Model. Simul. **8**, 803–835 (2010)
512. Pollard, T.D., Borisy, G.G.: Cellular motility driven by assembly and disassembly of actin filaments. Cell **112**, 453–465 (2003)
513. Popkov, V., Rakos, A., Williams, R.D., Kolomesisky, A.B., Schutz, G.M.: Localization of shocks in driven diffusive systems without particle number conservation. Phys. Rev. E **67**, 066117 (2003)
514. Posta, F., D'Orsogna, M.R., Chou, T.: Enhancement of cargo processivity by cooperating molecular motors. Phys. Chem. Chem. Phys. **11**, 4851–4860 (2009)
515. Pronina, E., Kolomeisky, A.: Spontaneous symmetry breaking in two-channel asymmetric exclusion processes with narrow entrances. J. Phys. A **40**, 2275–2287 (2007)
516. Prost, J., Chauwin, J.F., Peliti, L., Ajdari, A.: Asymmetric pumping of particles. Phys. Rev. Lett. **72**, 2652–2655 (1994)
517. Purcell, E.M.: Life at low reynolds number. Am. J. Phys. **45**, 3–11 (1977)
518. Qian, H.: Cooperativity in cellular biochemical processes. Annu. Rev. Biophys. **41**, 179–204 (2012)
519. Rafelski, S.M., Marshall, W.F.: Building the cell: design principles of cellular architecture. Mol. Cell Biol. **9**, 593–603 (2008)
520. Rafelski, S.M., Theriot, J.A.: Crawling toward a unified model of cell motility: spatial and temporal regulation of actin dynamics. Annu. Rev. Biochem. **73**, 209–239 (2004)
521. Raj, A., van Oudenaarden, A.: Nature, nurture, or chance: stochastic gene expression and its consequences Cell **135**, 216–226 (2008)
522. Redner, S.: Survival probability in a random velocity field. Phys. Rev. E **56**, 4967–4972 (1997)
523. Redner, S.: A Guide to First-Passage Processes. Cambridge University Press, Cambridge (2001)
524. Reed, M.C., Venakides, S., Blum, J.J.: Approximate traveling waves in linear reaction-hyperbolic equations. SIAM J. Appl. Math. **50**, 167–180 (1990)
525. Reese, L., Melbinger, A., Frey, E.: Crowding of molecular motors determines microtubule depolymerization Biophys. J. **101**, 2190–2200 (2011)
526. Reguera, D., Rubi, J.M.: Kinetic equations for diffusion in the presence of entropic barriers. Phys. Rev. E **64**, 061106 (2001)
527. Reguera, D., Schmid, G., Burada, P.S., Rubi, J.M., Reimann, P., Hanggi, P.: Entropic transport: kinetics, scaling, and control mechanisms. Phys. Rev. Lett. **96**, 130603 (2006)
528. Reichenbach, T., Franosch, T., Frey, E.: Exclusion processes with internal states. Phys. Rev. Lett. **97**, 050603 (2006)
529. Reimann, P.: Brownian motors: noisy transport far from equilibrium. Phys. Rep. **361**, 57–265 (2002)

530. Reimann, P., van den Broeck, C., Linke, H., Hanggi, P., Rubi, J.M., Perez-Madrid, A.: Diffusion in tilted periodic potentials: enhancement, universality and scaling. Phys. Rev. E **65**, 031104 (2002)
531. Reingruber, J., Holcman, D.: Narrow escape for a stochastically gated Brownian ligand. J. Phys. Condens. Matter **22**, 065103 (2010)
532. Reits, E.A., Neefjes, J.J.: From fixed to FRAP: measuring protein mobility and activity in living cells. Nat. Cell Biol. **3**, E145–E147 (2001)
533. Ribbeck, K., Gorlich, D.: Kinetic analysis of translocation through nuclear pore complexes. EMBO J. **21**, 2664–2671 (2001)
534. Ribbeck, K., Gorlich, D.: The permeability barrier of nuclear pore complexes appears to operate via hydrophobic exclusion. EMBO J. **21**, 2664–2671 (2002)
535. Ribiero, A.S., Smolander, O.P., Rajala, T., Hakkinen, A., Yli-Harja, O.: Delayed stochastic model of transcription at the single nucleotide level. J. Comput. Biol. **16**, 539–353 (2009)
536. Rice, S.A.: Diffusion-Limited Reactions. Elsevier, Amsterdam (1985)
537. Richter, P.H., Eigen, M.: Diffusion-controlled reaction rates in spheroidal geometry: application to repressor-operator association and membrane bound enzymes. Biophys. Chem. **2**, 255–263 (1974)
538. Rieder, C.L., Salmon, E.D.: Motile kinetochores and polar ejection forces dictate chromosome position on the vertebrate mitotic spindle. J. Cell Biol. **124**, 223–233 (1994)
539. Rietkerk, M., van de Koppel, J.: Regular pattern formation in real ecosystems. Trends. Ecol. Evol. **23**, 169–175 (2007)
540. Riggs, A.D., Bourgeois, S., Cohn, M.: The lac repressor-operator interaction iii: kinetic studies. J. Mol. Biol. **53**, 401–417 (1974)
541. Robbins, J.R., Monack, D., McCallum, S.J., Vegas, A., Pham, E., et al.: The making of a gradient: IcsA (VirG) polarity in Shigella flexneri. Mol. Microbiol. **41**, 861–872 (2001)
542. Rodenbeck, C., Karger, J., Hahn, K.: Calculating exact propagators in single-file systems via the reflection principle. Phys. Rev. E **57**, 4382–4397 (1998)
543. Romero, J.M., Gonzalez-Gaxiola, O., Chacon-Acosta, G.: Exact solutions to Fick-Jacobs equation. Int. J. Pure Appl. Math. **82**, 41–52 (2013)
544. Rook, M.S., Lu, M., Kosik, K.S.: CamKIIα 3' untranslated regions-directed mRNA translocation in living neurons: visualization by GFP linkage. J. Neurosci. **20**, 6385–6393 (2000)
545. Rout, M.P., Aitchison, J.D., Magnasco, M.O., Chait, B.T.: Virtual gating and nuclear transport: the hole picture. Trends Cell Biol. **13**, 622–628 (2003)
546. Roux, B., Allen, T., Berneche, S., Im, W.: Theoretical and computational models of biological ion channels. Q. Rev. Biophys. **37**, 15–103 (2004)
547. Rubi, J.M., Reguera, D.: Thermodynamics and stochastic dynamics of transport in confined media. Chem. Phys. **375**, 518–522 (2010)
548. Rubin, J., Terman, D.: Geometric singular perturbation analysis of neuronal dynamics. In: Fiedler, B. (ed.) Handbook of Dynamical Systems: Towards Applications, vol. 2. Elsevier, Amsterdam (2002)
549. Rzadzinska, A.K., Schneider, M.E., Davies, C., Riordan, G.P., Kachar, B.: An actin molecular treadmill and myosins maintain stereocilia functional architecture and self-renewal. J. Cell Biol. **164**, 887–897 (2004)
550. Sabatini, B.L., Maravall, M., Svoboda, K.: Ca^{2+} signaling in dendritic spines. Curr. Opin. Neurobiol. **11**, 349–356 (2001)
551. Saffman, P., Delbruck, M.: Brownian motion in biological membranes. Proc. Natl. Acad. Sci. USA **72**, 3111–3113 (1975)
552. Sahoo, S., Klumpp, S.: Backtracking dynamics of RNA polymerase: pausing and error correction. J. Phys. Condens. Matter **25**, 374104 (2013)
553. Salman, H., Abu-Arish, A., Oliel, S., Loyter, A., Klafter, J., Granek, R., Elbaum, M.: Nuclear localization signal peptides induce molecular delivery along microtubules. Biophys. J. **89**, 2134–2145 (2005)
554. Salsa, S.: Partial Differential Equations in Action. Springer, New York (2009)

555. Sanchez, A., Choubey, S., Kondev, J.: Regulation of noise in gene expression. Annu. Rev. Biophys. **42**, 469–491 (2013)
556. Santamaria, F., Wils, S., de Schutter, E., Augustine, G.J.: Anomalous diffusion in Purkinje cell dendrites caused by spines. Neuron **52**, 635–648 (2006)
557. Santillan, M., Mackey, M.C.: Influence of catabolite repression and inducer exclusion on the bistable behavior of the lac operon. Biophys. J. **86**, 1282–1292 (2004)
558. Sasai, M., Wolynes, P. G.: Stochastic gene expression as a many-body problem. Proc. Natl. Acad. Sci. **100**, 2374–2379 (2003)
559. Saunders, T.E., Pan, K.Z., Angel, A., Guan, Y., Shah, J.V., Howard, M., Chang, F.: Noise reduction in the intracellular pom1p gradient by a dynamic clustering mechanism. Dev. Cell **22**, 558–572 (2012)
560. Savage, N.S., Layton, A.T., Lew, D.J.: Mechanistic model of polarity in yeast. Mol. Biol. Cell **23**, 1998–2013 (2012)
561. Saxton, M.J.: The membrane skeleton of erythrocytes: a percolation model. I erythrocytes: a percolation analysis. Biophys. J. **57**, 1167–1177 (1990)
562. Saxton, M.J.: Anomalous diffusion due to obstacles: a monte carlo study. Biophys. J. **66**, 394–401 (1994)
563. Saxton, M.J.: Single particle tracking: effects of corrals. Biophys. J. **69**, 389–398 (1995)
564. Saxton, M.J.: Anomalous diffusion due to binding: a Monte Carlo study. Biophys. J. **70**, 1250–1262 (1996)
565. Saxton, M.J.: A biological interpretation of transient anomalous subdiffusion. I: qualitative model. Biophys. J. **92**, 1178–1191 (2007)
566. Saxton, M.J., Jacobson, K.: Single-particle tracking: applications to membrane dynamics. Annu. Rev. Biophys. Biomol. Struct. **26**, 373–399 (1997)
567. Schadschneider, A., Chowdhury, D., Nishinari, K.: Stochastic transport in complex systems: from molecules to vehicles. Elsevier, Amsterdam (2010)
568. Schaus, T.E., Borisy, G.G.: Performance of a population of independent filaments in lamellipodial protrusion. Biophys. J. **95**, 1393–1411(2008)
569. Scher, H., Montroll, E.W.: Anomalous transit-time dispersion in amorphous solids. Phys. Rev. B **12**, 2455–2477 (1975)
570. Schnapp, B.J., Gelles, J., Sheetz, M.P.: Nanometer-scale measurements using video light microscopy. Cell Motil. Cytoskeleton **10**, 47–53 (1988)
571. Schnitzer, M., Visscher, K., Block, S.: Force production by single kinesin motors. Nat. Cell Biol. **2**, 718–723 (2000)
572. Scholey, J.M.: Intraflagellar transport. Annu. Rev. Cell Dev. Biol. **19**, 423–443 (2003)
573. Schumacher, L.J., Woolley, T.E., Baker, R.E.: Noise-induced temporal dynamics in Turing systems. Phys. Rev. E **87**, 042719 (2013)
574. Schuss, Z.: Theory and applications of stochastic processes: an analytical approach. In: Applied Mathematical Sciences, vol. 170. Springer, New York (2010)
575. Schuss, Z., Nadller, B., Eisenberg, R.S. : Derivation of PNP equations in bath and channel from a molecular model. Phys. Rev. E **64**, 036116 (2001)
576. Schuss, Z., Singer, A., Holcman, D.: The narrow escape problem for diffusion in cellular microdomains. Proc. Natl. Acad. Sci. USA **104**, 16098–16103 (2007)
577. Schuster, M., Lipowsky, R., Assmann, M.A., Lenz, P., Steinberg, G.: Transient binding of dynein controls bidirectional long-range motility of early endosomes. Proc. Natl Acad. Sci. USA **108**, 3618–3623 (2011)
578. Scott, M., Ingalls, B., Kearns, M.: Estimation of intrinsic and extrinsic noise in models of nonlinear genetic networks. Chaos **16**, 026107 (2006)
579. Segel, L.A., Edelstein-Keshet, L.: A Primer on Mathematical Models in Biology. SIAM, Philadelphia (2013)
580. Sekimoto, K., Triller, A.: Compatibility between itinerant synaptic receptors and stable postsynaptic structure. Phys. Rev. E **79**, 031905 (2009)
581. Semplice, M., Veglio, A., Naldi, G., Serini, G., Gamba, A.: A bistable model of cell polarity. PLoS One **7**, e30977 (2012)

582. Shannon, C.E., Weaver, W.: The Mathematical Theory of Communication. University of Illinois Press, Urbana (1949)
583. Shapere, A., Wilczek, F.: Geometry of self-propulsion at low Reynolds number. J. Fluid Mech. **198**, 557–585 (1989)
584. Shaw, L.B., Zia, R.K.P., Lee, K.H.: Totally asymmetric exclusion process with extended objects: a model for protein synthesis. Phys. Rev. E **68**, 021910 (2003)
585. Sheinman, M., Benichou, O., Kafri, Y., Voituriez, R.: Classes of fast and specific search mechanisms for proteins on DNA. Rep. Prog. Phys. **75**, 026601 (2012)
586. Sherman, A., Rinzel, R., Keizer, J.: Emergence of organized bursting in clusters of pancreatic beta-cells by channel sharing. Biophys. J. **54**, 411–425 (1988)
587. Shouval, H.Z.: Clusters of interacting receptors can stabilize synaptic efficacies. Proc. Natl. Acad Sci. USA **102**, 14440–14445 (2005)
588. Shtylla, B., Keener, J.P.: A mechanomolecular model for the movement of chromosomes during mitosis driven by a minimal kinetochore bicyclic cascade. J. Theor. Biol. **263**, 455–70 (2010)
589. Shtylla, B., Keener, J.P.: A mathematical model for force generation at the kinetochore-microtubule interface. SIAM J. Appl. Math. **71**, 1821–1848 (2011)
590. Shuai, J.W., Jung, P.: Stochastic properties of Ca^{2+} release of Inositol 1,4,5-Triphosphate receptor clusters. Biophys. J. **83**, 87–97 (2002)
591. Shvartsman, S.Y., Baker, R.E.: Mathematical models of morphogen gradients and their effects on gene expression. Rev. Dev. Biol. **1**, 715–730 (2012)
592. Simon, S.M., Peskin, C.P., Oster, G.F.: What drives the translocation of proteins? Proc. Natl. Acad. Sci. USA **89**, 3770–3774 (1992)
593. Singer, S.J., Nicolson, G.L.: The fluid mosaic model of the structure of the cell membrane. Science **175**, 720–731 (1972)
594. Singer, A., Schuss, Z., Holcman, D.: Narrow escape, Part II: the circular disc. J. Stat. Phys. **122**, 465–489 (2006)
595. Singer, A., Schuss, Z., Holcman, D.: Narrow escape, Part III: nonsmooth domains and Riemann surfaces. J. Stat. Phys. **122**, 491–509 (2006)
596. Skibbens, R.V., Petrie-Skeen,V., Salmon, E.D.: Directional instability of kinetochore motility during chromosome congression and segregation in mitotic newt lung cells: a push pull mechanism. J. Cell Biol. **122**, 859–875 (1993)
597. Slaughter, B.D., Das, A., Schwartz, J.W., Rubinstein, B., Li, R.: Dual modes of Cdc42 recycling fine-tune polarized morphogenesis. Dev. Cell **17**, 823–835 (2009)
598. Slaughter, B.D., Smith, S.E., Li, R.: Symmetry breaking in the life cycle of budding yeast. Cold Spring Harb. Perspect. Biol. **1**, a003384 (2009)
599. Smiley, M.W., Proulx, S.R.: Gene expression dynamics in randomly varying environments. J. Math. Biol. **61**, 231–251 (2010)
600. Smith, G.D.: Modeling the stochastic gating of ion channels, chap. 11. In: Fall, C., Marland, E.S., Wagner, J.M., Tyson, J.J. (eds.) Computational Cell Biology. Springer, New York (2002)
601. Smith, D.A., Simmons, R.M.: Models of motor-assisted transport of intracellular particles. Biophys. J. **80**, 45–68 (2001)
602. Smolen, P., Baxter, D.A., Byrne, J.H.: Modeling circadian oscillations with interlocking positive and negative feedback loops. J. Neurosci. **21**, 6644–6656 (2001)
603. Smoluchowski, M.V.: Z. Phys. Chem. **92**, 129–168 (1917)
604. Sorra, K.E., Harris, K.M.: Overview on the structure, composition, function, development, and plasticity of hippocampal dendritic spines. Hippocampus **10**, 501–511 (2000)
605. Sourjik, V., Berg, H.C.: Receptor sensitivity in bacterial chemotaxis. Proc. Natl. Acad. Sci. USA **99**, 123–127 (2002)
606. Sprague, B.L., Pearson, C.G., Maddox, P.S., Bloom, K.S., Salmon, E.D., Odde, D.J.: Mechanisms of microtubule-based kinetochore positioning in the yeast metaphase spindle. Biophys. J. **84**, 1–18 (2003)

607. Stauffer, E.A., Scarborough, J.D., Hirono, M., Miller, E.D., Shah, K., Mercer, J.A., Holt, J.R., Gillespie, P.G.: Fast adaptation in vestibular hair cells requires myosin-1c activity. Neuron **47**, 541–553 (2005)
608. Stephens, R.E.: Quantal tektin synthesis and ciliary length in sea-urchin embryos. J. Cell Sci. **92**, 403–413 (1989)
609. Stephens, D.J., Pepperkok, R.: Illuminating the secretory pathway: when do we need vesicles? J. Cell Sci. **114**, 1053–1059 (2001)
610. Steyger, P.S., Gillespie, P.G., Baird, R.A.: Myosin Ib is located at tip link anchors in vestibular hair bundles. J. Neurosci. **18**, 4603–4615 (1998)
611. Stolarska, M.A., Kim, Y., Othmer, H.G.: Multi-scale models of cell and tissue dynamics. Phil Trans. Roy. Soc. A **367**, 3525–3553 (2009)
612. Stratonovich, R.L.: Radiotekhnika i Elektronika **3**, 497–511 (1958)
613. Straube, R., Ward, M.J., Falcke, M.: Reaction rate of small diffusing molecules on a cylindrical membrane. J. Stat. Phys. **129** 377–405 (2007)
614. Stukalin, E.B., Kolomeisky, A.B.: Simple growth models of rigid multifilament biopolymers. J. Chem. Phys. **121**, 1097–1104 (2004)
615. Sung, W., Park, P.J.: Polymer translocation through a pore in a membrane. Phys. Rev. Lett. **77**, 783–786 (1996)
616. Sung, B.J., Yethiraj, A.: Lateral diffusion of proteins in the plasma membrane: spatial tessellation and percolation theory. J. Phys. Chem. B **112**, 143–149 (2008)
617. Sutradhar, S., Paul, R.: Tug-of-war between opposing molecular motors explains chromosomal oscillation during mitosis. J. Theor. Biol. **334**, 56–69 (2014)
618. Swillens, S., Dupont, G., Combettes, L., Champeil, P.: From calcium blips to calcium puffs: theoretical analysis of the requirement for interchannel communication. Proc. Natl. Acad. Sci. (USA) **96**, 13750–13755 (1999)
619. Szymanski, J., Weiss, M.: Elucidating the origin of anomalous subdiffusion in crowded fluids. Phys. Rev. Lett. **103**, 038102 (2009)
620. Taloni, A., Lomholt, M.A.: Langevin formulation for single-file diffusion. Phys. Rev. E **78**, 051116 (2008)
621. Taloni, A., Marchesoni, F.: Single-file diffusion on a periodic substrate. Phys. Rev. Lett. **96**, 020601 (2006)
622. Telley, I.A., Bieling, P., Surrey, T.: Obstacles on the microtubule reduce the processivity of Kinesin-1 in a minimal in vitro system and in cell extract. Biophys. J. **96**, 3341–3353 (2009)
623. Thanbichler, M., Shapiro, L.: MipZ, a spatial regulator coordinating chromosome segregation with cell division in Caulobacter. Cell **126**, 147–162 (2006)
624. Thattai, M., van Oudenaarden, A.: Attenuation of noise in ultrasensitive signaling cascades. Biophys. J. **82**, 2943–2950 (2001)
625. Thattai, M., van Oudenaarden, A.: Intrinsic noise in gene regulatory networks. Proc. Natl. Acad. Sci. USA **98**, 8614–8619 (2001)
626. Theriot, J.A., Mitchison, T.J., Tilney, L.G., Portnoy, D.A.: The rate of actin-based motility of intracellular Listeria monocytogenes equals the rate of actin polymerization. Nature. **357**, 257–260 (1992)
627. Tilney, L.G., Portnoy, D.A.: Actin filaments and the growth, movement, and spread of the intracellular bacterial parasite, Listeria monocytogenes. J. Cell Biol. **109**, 1597–1608 (1989)
628. Tindall, M.J., Porter, S.L., Maini, P.K., Gaglia, G., Armitage, J.P.: Overview of mathematical approaches to model bacterial chemotaxis I: the single cell. Bull. Math. Biol. **70**, 1525–1569 (2008)
629. Tischer, C., ten Wolde, P.R., Dogterom, M.: Providing positional information with active transport on dynamic microtubules. Biophys. J. **99**, 726–35 (2010)
630. Tkacik, G., Bialek, W.: Diffusion, dimensionality, and noise in transcriptional regulation. Phys. Rev. E **79**, 051901 (2009)
631. Tkacik, G., Walczak, A.M.: Information transmission in genetic regulatory networks: a review. J. Phys. Condens. Matter **23**, 153102 (2011)

632. Tkacik, G., Callan, C.G. Jr., Bialek, W.: Information flow and optimization in transcriptional regulation. Proc. Natl. Acad. Sci. USA **105**, 12265–12270 (2008)
633. Tkacik, G., Gregor, T., Bialek, W.: The Role of Input Noise in Transcriptional Regulation. PLoS One **3**, e2774 (2008)
634. Tkacik, G., Walczak, A.M., Bialek, W.: Optimizing information flow in small genetic networks. Phys. Rev. E **80**, 031920 (2009)
635. Tkacik, G., Walczak, A.M., Bialek, W.: Optimizing information flow in small genetic networks III. A self-interacting gene. Phys. Rev. E **85**, 041903 (2012)
636. Toomre, D., Keller, P., White, J., Olivo, J.C., Simons, K.: Dual-color visualization of trans-Golgi network to plasma membrane traffic along microtubules in living cells. J. Cell Sci. **112** 21–33 (1999)
637. Tostevin, F.: Precision of sensing cell length via concentration gradients. Biophys. J. **100**, 294–303 (2011)
638. Tostevin, F., ten Wolde, P.R.: Mutual information between input and output trajectories of biochemical networks. Phys. Rev. Lett. **102**, 218101 (2011)
639. Tostevin, F., Rein ten Wolde, P., Howard, M.: Fundamental limits to position determination by concentration gradients. PLoS Comput. Biol. **3**, e78 (2007)
640. Touchette, H.: The large deviation approach to statistical mechanics. Phys. Rep. **478**, 1–69 (2009)
641. Tran, E.J., Wente, S.R.: Dynamic nuclear pore complexes: life on the edge. Cell **125**, 1041–1053 (2006)
642. Triller, A., Choquet, D.: Surface trafficking of receptors between synaptic and extrasynaptic membranes: and yet they do move! Trends Neurosci. **28**, 133–139 (2005)
643. Triller, A., Choquet, D.: New concepts in synaptic biology derived from single-molecule imaging. Neuron **59**, 359–374 (2008)
644. Tsimiring, L.S.: Noise in biology. Rep. Prog. Phys. **77**, 026601 (2014)
645. Tu, Y.: Quantitative modeling of bacterial chemotaxis: signal amplification and accurate adaptation. Annu. Rev. Biophys. **42**, 337–359 (2013)
646. Turing, A.M.: The chemical basis of morphogenesis. Philos. Trans. Roy. Soc. Lond. B **237**, 37–72 (1952)
647. Turner, T.E., Schnell, S., Burrage, K.: Stochastic approaches for modelling in vivo reactions. Comput. Biol. Chem. **28**, 165–178 (2004)
648. Tyson, J.J., Hong, C.I., Thron, C.D., Novak, B.: A simple model of circadian rhythms based on dimerization and proteolysis of PER and TIM. Biophys. J. **77**, 2411–2417 (1999)
649. Vale, R.D., Funatsu, T., Pierce, D.W., Romberg, L., Harad, Y., Yanagida, T.: Direct observation of single kinesin molecules moving along microtubules. Nature **380**, 451–453 (1996)
650. Vance, W., Ross, J.: Fluctuations near limit cycles in chemical reaction systems. J. Chem. Phys. **105**, 479–487 (1996)
651. van Kampen, N.G.: Stochastic Processes in Physics and Chemistry. North-Holland, Amsterdam (1992)
652. Varga, V., et al.: Yeast kinesin-8 depolymerizes microtubules in a length-dependent manner. Nat. Cell Biol. **8**, 957–962 (2006)
653. Vellela, M., Qian, H.: Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the Schlögl model revisited. J. Roy. Soc. Interface **6**, 925–940 (2009)
654. Vereb, G., Szollosi, J., Nagy, J.M.P., Farkas, T.: Dynamic, yet structured: the cell membrane three decades after the Singer-Nicolson model. Proc. Natl. Acad. Sci. USA **100**, 8053–8058 (2003)
655. Vershinin, M., Carter, B.C., Razafsky, D.S., King, S.J., Gross, S.P.: Multiple-motor based transport and its regulation by Tau. Proc. Natl. Acad. Sci. USA **104**, 87–92 (2007)
656. Vicente-Manzanares, M., Choi, C.K., Horwitz, A.R.: Integrins in cell migration—the actin connection. J. Cell Sci. **122**, 199–206 (2009)
657. Vilfan, A., Duke, T.: Two adaptation processes in auditory hair cells together can provide an active amplifier. Biophys. J. **85**, 191–203 (2003)

658. Visscher, K., Schnitzer, M., Block, S.: Single kinesin molecules studied with a molecular force clamp. Nature **400**, 184–189 (1999)
659. Viswanathan, G., Buldyrev, S., Havlin, S., da Luz, M., Raposo, E., Stanley, H.: Optimizing the success of random searches. Nature **401**, 911–914 (1999)
660. Viswanathan, G., Bartumeus, F., Buldyrev, S., Catalan, J., Fulco, U., Havlin, S., da Luz, M., Lyra, M., Raposo, E., Stanley, H.: Levy flight random searches in biological phenomena. Physica A **314**, 208–213 (2002)
661. Voliotis, M., Cohen, N., Molina-Paris, C., Liverpool, T.: Fluctuations, pauses, and backtracking in DNA transcription. Biophys. J. **94**, 334–348 (2008)
662. Vorotnikov, D.: Analytical aspects of the Brownian motor effect in randomly flashing ratchets. J. Math. Biol. 68:1677–1705 (2014)
663. Wainrib, G., Thieullen, M., Pakdaman, K.: Reduction of stochastic conductance-based neuron models with time-scales separation. J. Comput. Neurosci. **32**, 327–46 (2012)
664. Walczak, A. M., Sasai, M., Wolynes, P. G.: Self-consistent proteomic field theory of stochastic gene switches. Biophys. J. **88**, 828–850 (2005)
665. Walczak, A.M., Tkacik, G., Bialek, W.: Optimizing information flow in small genetic networks. II. Feed-forward interactions. Phys. Rev. E **81**, 041905 (2010)
666. Walgraef, D.: Spatio-Temporal Pattern Formation. Springer, New York (1997)
667. Walther, G.R., Maree, A.F., Edelstein-Keshet, L., Grieneisen, V.A.: Deterministic versus stochastic cell polarisation through wave-pinning. Bull. Math. Biol. **74**, 2570–2599 (2012)
668. Wang, K.G.: Long-time correlation effects and biased anomalous diffusion. Phys. Rev. A **45**, 833–837 (1992)
669. Wang, H., Elston, T.C., Mogilner, A., Oster, G.: Force generation in RNA polymerase. Biophys. J. **74**, 1186–1202 (1998)
670. Wang, S.Q., Song, L.S., Xu, L., Meissner, G., Lakatta, E.G., Rios, E., Stern, M.D., Cheng, H.: Thermodynamically irreversible gating of ryanodine receptors in situ revealed by stereotyped duration of release of Ca21 sparks. Biophys. J. **83**, 242–251 (2002)
671. Ward, M.J., Henshaw, W.D., Keller, J.B.: Summing logarithmic expansions for singularly perturbed eigenvalue problems. SIAM J. Appl. Math. **53**, 799–828 (1993)
672. Wartlick, O., Kicheva, A., Gonzalez-Gaitan, M.: Morphogen gradient formation. Cold Spring Harb. Perspect. Biol. **1**, a001255 (2009)
673. Weber, S.C., Spakowitz, A.J., Theriot, J.A.: Bacterial chromosomal loci move subdiffusively through a viscoelastic cytoplasm. Phys. Rev. Lett. **104**, 238102 (2010)
674. Wedlich-Soldner, R., Wai, S.C., Schmidt, T., Li, R.: Robust cell polarity is a dynamic state established by coupling transport and GTPase signaling. J. Cell. Biol. **166**, 889–900 (2004)
675. Weigel, A.V., Simon, B., Tamkun, M.M., Krapf, D.: Ergodic and nonergodic processes coexist in the plasma membrane as observed by single-molecule tracking. Proc. Natl. Acad. Sci. USA **108**, 6438–6443 (2011)
676. Welf, E.S., Johnson, H.E., Haugh, J.M.: Bidirectional coupling between integrin-mediated signaling and actomyosin mechanics explains matrix-dependent intermittency of leading-edge motility. Mol. Biol. Cell **24**, 3945–3955 (2013)
677. Welte, M.A.: Bidirectional transport along microtubules. Curr. Biol. **14**, R525–R537 (2004)
678. White, J.A., Rubinstein, J.T., Kay, A.R.: Channel noise in neurons. Trends Neurosci. **23**, 131–137 (2000)
679. Whittaker, G.R., Kann, M., Helenius, A.: Viral entry into the nucleus. Annu. Rev. Cell Dev. Biol. **16**, 627–651 (2000)
680. Wiley, H.S., Shvartsman, S.Y., Lauffenburger, D.A.: Computational modeling of the EGF-receptor system: a paradigm for systems biology. Trends Cell Biol. **13**, 43–50 (2003)
681. Williams, D.: Probability with Martingales. Cambridge University Press, Cambridge (1991)
682. Williams, G.S.B., Huertas, M.A., Sobie, E.A., Jafri, M.S., Smith, G.D.: A probability density approach to modeling local control of calcium-induced calcium release in cardiac myocytes. Biophys. J. **92**, 2311–2328 (2007)

683. Williams, G.S.B., Huertas, M.A., Sobie, E.A., Jafri, M.S., Smith, G.D.: Moment closure for local control models of calcium-induced calcium release in cardiac myocytes. Biophys. J. **95**, 1689–1703 (2008)
684. Winter, R.B., von Hippel, P.H.: Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. Biochemistry **20**, 6929–6948 (1981)
685. Wollman, R., Cytynbaum, E.N., Jones, J.T., Meyer, T., Scholey, J.M., Mogilner, A.: Efficient chromosome capture requires a bias in the 'search-and-capture' process during mitotic-spindle assembly. Curr. Biol. **15**, 828–832 (2005)
686. Wolpert, L.: Positional information and the spatial pattern of cellular differentiation. J. Theor. Biol. **25**, 1–47 (1969)
687. Wolpert, L.: Principles of Development. Oxford University Press, Oxford (2006)
688. Wong, P., Gladney, S., Keasling, J.D.: Mathematical model of the lac operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. Biotech. Prog. **13**, 132–143 (1997)
689. Woolley, T.E., Baker, R.E., Gaffney, E.A., Maini, P.K.: Stochastic reaction and diffusion on growing domains: understanding the breakdown of robust pattern formation. Phys. Rev. E **84**, 046216 (2011)
690. Yamada, Y., Peskin, C.: A look-ahead model for the elongation dynamics of transcription. Biophys. J. **96**, 3015–3031 (2009)
691. Yang, Y.M., Austin, R.H., Cox, E.C.: Single molecule measurements of repressor protein 1D diffusion on DNA. Phys. Rev. Lett. **97**, 048302 (2006)
692. Yao, Y., Choi, J., Parker, I.: Quantal puff of intracellular Ca^{2+} evoked by inositol triphosphate in Xenopus oocytes. J. Physiol. **482**, 533–553 (1995)
693. Yildirim, N., Mackey, M.C.: Feedback regulation in the lactose operon: a mathematical modeling study and comparison with experimental data. Biophys. J. **84**, 2841–2851 (2003)
694. Yildirim, N., Santillan, M., Horike, D., Mackey, M.C.: Dynamics and bistability in a reduced model of the lac operon. Chaos **14**, 279–292 (2004)
695. Yildiz, A., Tomishige, M., Vale, R.D., Selvin, P.R.: Kinesin walks hand-over-hand. Science **303**, 676–678 (2004)
696. Young, G.W.D., Keizer, J.: A single pool IP_3-receptor model for agonist stimulated Ca^{2+} oscillations. Proc. Natl. Acad. Sci. (USA) **89**, 9895–9899 (1992)
697. Yu, J., Moffitt, J., Hetherington, C.L., Bustamante, C., Oster, G.: Mechanochemistry of a viral DNA packaging motor. J. Mol. Biol. **400**, 186–203 (2010)
698. Yuste, R., Majewska, A., Holtho, K.: From form to function: calcium compartmentalization in dendritic spines. Nat. Neurosci. **3**, 653–659 (2000)
699. Zahradnikova, A., Zahradnik, I.: A minimal gating model for the cardiac Ca^{2+} release channel. Biophys. J. **71**, 2996–3012 (1996)
700. Zaikin, A.N., Zhabotinsky, A.M.: Concentration wave propagation in two-dimensional liquid-phase self-oscillating system. Nature **225**, 535–537 (1970)
701. Zhang, X.-J., Qian, H., Qian, M.: Stochastic theory of nonequilibrium steady states and its applications. Part I. Phys. Rep. **510**, 1–89 (2012)
702. Zhou, H.X.: A model for the mediation of processivity of DNA-targeting proteins by nonspecific binding: dependence on DNA length and presence of obstacles. Biophys. J. **88**, 1608–1615 (2005)
703. Zia, R.K.P., Dong, J.J., Schmittmann, B.: Modeling translation in protein synthesis with TASEP: A tutorial and recent developments. J. Stat. Phys. **144**, 405–428 (2011)
704. Zilman, A., Talia, S.D., Chait, B.T., Rout, M.P., Magnasco, M.O.: Efficiency, selectivity and robustness of nucleocytoplasmic transport. PLoS Comput. Biol. **3**, e125 (2007)
705. Ziv, E., Nemenman, I., Wiggins, C.H.: Optimal signal processing in small stochastic biochemical networks. PLoS One **2**, e1077 (2007)
706. Zumofen, G., Klafter, J., Blumen, A.: Enhanced diffusion in random velocity fields. Phys. Rev. A **42**, 4601–4608 (1990)
707. Zwanzig, R.: Rate processes with dynamical disorder. Acc. Chem. Res. **23**, 148–152 (1990)
708. Zwanzig, R.: Diffusion past an entropy barrier. J. Phys. Chem. **96**, 3926–3930 (1992)

Index

P.C. Bressloff, *Stochastic Processes in Cell Biology*, Interdisciplinary Applied Mathematics 41, DOI 10.1007/978-3-319-08488-6