Molecules in motion: a theoretical study of noise in gene expression and cell signaling
Dobrzynski, M.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 1

Introduction

Contents

1.1 Stochasticity fundamentals .............................................. 4
1.2 First example .......................................................... 11
  1.2.1 Physiological implications of bursts .......................... 13
1.3 Timing of biochemical reactions ................................. 17
  1.3.1 Sequential processes ........................................... 17
  1.3.2 Diffusion-limited reactions ................................... 20
1.4 Organization of the thesis ........................................... 21

Throw a pebble and... a cat from the window and you will hear just one thump in
the courtyard, possibly preceded by a squeaking meow. Theoretically. The air friction will
most likely slow down the furry pet a little while it desperately attempts to adjust the
position before the inevitable. Prior to discovering the celebrated relation for the motion
of a mass in a gravitational field one needs to consider a highly abstracted version of the
problem. An easy procedure for an apt high school student in the XXI century, but a
real challenge for those unaware of Newton’s *Principia*. Turning the stone and the cat
into point masses and letting them enjoy the fall in a vacuum, should do the trick and
reveal that neither the object’s mass, nor its shape affect the motion, the fact expressed
by Galileo nearly 400 years ago: *All objects fall at the same rate in a vacuum.*

There are at least three conclusions to infer from this example. First, an animal, though
very different in our perception from a dull stone, belongs to the same realm where the
fundamental laws of physics apply. Quantum theory, currently the most fundamental
description of matter, is as capable of explaining the formation of a covalent bonding of
carbon atoms in a diamond lattice as the properties of a hydrogen bond to which nucleic
acids and proteins owe their three-dimensional structure.

Secondly, questions about physical processes can be successfully addressed even with
theories that inherently have limited scopes. While certainly required for describing phe-
nomena on atomic scales, quantum mechanics (QM) looses its appeal when applied to a
large collection of particles such as a stone or a cat. Although an architect designing a
bridge might indirectly profit from QM through advancements in material science, he will
certainly succeed without being aware of all the intricacies of the Schrödinger equation
and the Hilbert space. Classical Newtonian mechanics, laws of which follow from the laws
of QM in the limit of macroscopic systems, is sufficient to describe tensions and loads of
a massive steel construction and the falling of cats.

The final observation regards inferring the underlying theoretical framework. When
skimming through an introductory physics course book, students often come to a (hope-
fully superficial) conclusion that physics has actually very little to do with the real world.
Point masses, frictionless motion, approximations even to seemingly simple problems (physical pendulum, three-body problem) do not have an immediate counterpart in reality. Nonetheless, a significant idealization of a problem eases the discovery of fundamental laws. A movement of a block involves friction, heat transfer, possible deformations. In most of the cases, however, a good approximation can be obtained by omitting those effects and trading them for frictionless dynamics of point masses. Naturally, a model is realistic as long as it provides testable predictions but more importantly the idealized description allows to discover a more general framework.

Formulation of classical mechanics using the Lagrangian formalism, although applicable to a narrow class of systems without an approximation, gives insights into the structure of the physics theory. For instance, the law of conservation of energy is a very deep and far-reaching principle present in such diverse physical phenomena as electromagnetic radiation, general relativity or quantum mechanics. A relatively straightforward algebra reveals another deep result – the principle of least action. In fact, laws of motion of particles in gravitational fields, laws of electricity and magnetism, motion of particles in electric fields, all have a common basis: the principle of least action. Laws of thermodynamics are simply statistical claims about large numbers of degrees of freedom, given that those degrees of freedom are governed by the underlying principle of least action.

Although numerous details are deemed unimportant, simple models give particularly valuable insights into the problem; especially if the mathematical representation is tractable enough to allow for an informative and closed solution. Unfortunately, the amount of analytically solvable models in physics is limited. The situation in biological sciences seems even grimmer and the reasons for this are the following. For generations natural sciences have been dominated by strong belief in the power of reductionism. Rightly to be so. A long struggle to discover the nature of forces holding matter together resulted in a solid theoretical framework followed by a surge of then unthinkable applications. Parallel to these ground-breaking discoveries biologists continued to be faced with a multitude of new species, forms, animal behaviors and habits. A profound contribution of Darwin cleared the picture, although being developed almost the same year as Maxwell’s equations (unfortunately) still continues to be the subject of many heated debates. Only dawn of molecular biology and genetics in the twentieth century shed new light on mechanisms reasoned a hundred years earlier. But then the problem became even more clear. The quest to reveal fundamental principles in biology has become marred by a staggering number of components constituting a living organism, the degree of interaction between the elements and finally the huge number of states exhibited by such an interacting network. Even a simple single-cell organism involves processes spanning many orders of magnitude: from nanosecond timing of structural motion of bio-polymers, to molecular clocks ticking at night-and-day intervals, or month or even year-long-lasting states of biochemical switches. All of these factors make it difficult to view the system as a collection of independent entities. But living organisms are subject to the laws of physics. Is it then possible to explain their design from the first principles using the same tools as in physics? Is it reasonable to assume that once we understand all the fundamental laws governing micro-world, the macroscopic description will follow [Binder 2009]?7

In order to survive in the environmental niche, even the smallest organism has to reliably pass genetic information to the progeny, a multitude of intracellular processes needs to be coordinated in order to proliferate, metabolize, respond to changes in nutrient con-
centration or to sudden appearance of toxic compounds. On the other hand, colonization or adaptation to new environmental conditions requires changes in the phenotype accessible only through genetic mutations, which give rise to new metabolic, sensory or physical features. Some of the survival strategies involve genotypes capable of generating various phenotypes optimal for different environments. Various regulatory mechanisms exploit noise to randomize outcomes where variability is advantageous [McAdams & Arkin 1999]. In order to achieve all of these goals, regulation of biochemical processes evolved such that the organism may profit from retaining stochasticity inherent in biochemical reactions, while other structures evolve to suppress the noise which would disrupt precision of cellular physiology.

Thermal fluctuations may easily alter the rates of chemical reactions, local concentrations of reactants or the composition of genetic code [Dronamraju 1999]. In a number of experiments, isogenic populations starting from the same initial conditions have been propagated into cells with entirely different molecular makeup [Elowitz et al. 2002, Blake et al. 2006, Spencer et al. 2009] or proliferated into phenotypically distinct cellular entities with diversified biological functions [Balaban et al. 2004, Feinerman et al. 2008, Chang et al. 2008] – a convincing demonstration of how random molecular events may affect the macroscopic observable: the phenotype.

If molecular fluctuations are so abundant, it is truly remarkable that seemingly fragile molecular structures inside living systems reliably hold genetic information even for thousands of years, perform uncountable enzymatic reactions, protect the organisms against the influence of harsh environments and give rise to countless life forms, shapes and behaviors. As might be expected, the composition (or design) of an organism successfully populating a niche reflects the two tendencies mentioned earlier: flexibility allowing for acquisition of new features and robustness facilitating reliable information processing. For instance, sensing changes in the nutrient concentration in the environment of a simple bacterial cell is prone to uncertainties due to thermal noise. Attaining a level of signal-to-noise ratio that would allow for appropriate and possibly rapid cellular response requires a specific architecture of molecular structures (a sensor) and corresponding biochemical reactions (a signaling and a regulatory network). On top of that, the number of biomolecules involved in the response has to be such that the network itself is not activated by a sudden fluctuation in the concentration of reactants.

Thus, recognizing how organisms may benefit from molecular fluctuations and when these stochastic effects are suppressed gives valuable insights into how fundamental physical constraints shape physiology of biological systems in the course of evolution. For instance, fluctuations are attenuated through the structure of biochemical networks (e.g. cascade architecture of eukaryotic signaling networks [Thattai & van Oudenaarden 2002, Hooshangi et al. 2005, Bruggeman et al. 2009]) or through properties of the stochastic system itself (e.g. sequential processes described in more detail further in this chapter). The evolution of structures capable of reducing molecular fluctuations confers fitness advantage for the organism. The design of these structures such as the number of levels in the cascade or steps in the sequence, the number of molecules involved in the process reflects the direction that the evolutionary process has taken in order to overcome physical constraints in processing of the extracellular signal or limitations in the speed of chemical reactions.

Manifestation of stochastic phenomena in cellular processes has some practical im-
plications for the modeling community too. Analysis of problems involving stochasticity requires a more complex theoretical and computational apparatus than deterministic processes. Therefore, it is of great importance to recognize deterministic regimes in biochemical processes. Typically, for a large number of molecules the significance of molecular fluctuations diminishes. The complexity of theoretical description reduces dramatically and so does the computational cost. Large time scale separation confers a similar advantage: very fast fluctuations may be averaged out by much slower downstream processes effectively alleviating the need for a detailed stochastic description of the fast module.

The exposition in the following sections will focus on stochastic phenomena in a single cell. To address that, we shall introduce the framework for the analysis of stochastic events, namely the waiting-time or the first-passage time theory. We will draw parallels to familiar macroscopic approaches by illustrating the limits of the stochastic theory for a large number of molecules. By analyzing the effect of various waiting-time statistics in protein synthesis and degradation we shall gain insights into the source of precise timing in stochastic systems as well as the reason why a deterministic approximation works so well in theoretical biology. In particular, we shall discuss the exponential regime of the waiting times which allows for significant simplification of sequential and diffusion-limited biochemical processes. The exponential approximation contributes to a largely unexplored topic of the interface between stochastic and deterministic theoretical modeling in biology.

1.1 Stochasticity fundamentals

Biochemical transitions typically of interest for molecular systems biology involve macromolecules sized above 5 nm; a so called mesoscopic level. Confront that with $\approx 0.3 \text{ nm}$ – the effective diameter of a water molecule or an ion. An *E. coli*, a model prokaryotic organism contains $2.5 \times 10^{10}$ (25 billion!) water molecules and as little as $\approx 4$ million proteins in a 2 $\mu$m-long ellipsoid cell. An enormous disparity of time scales of atomic interactions between small molecules (e.g. water, ions) and large bio-molecules (e.g. proteins, nucleic acid chains, membrane lipids) permits statistical claims about the effective forces exerted on the latter. Hence, the movement of biomolecules can be described much simpler by introducing a stochastic term accounting for random collisions of a protein, for instance, with much smaller molecules of the solvent. The complexity of the resulting equations of motion (Einstein-Smoluchowski diffusion equation) is reduced dramatically compared to the initial description accounting for the motion of all molecules, the solvent and the solute. If diffusive encounters of molecules are much faster than the time required to undergo a chemical transition, the spatial aspect of the problem can be discarded and an even simpler model materializes. The complication remains, however. Thermal fluctuations affecting the rates of chemical reactions remain in the form of probabilistic terms similar to those in the synthesis-degradation problem. The effect of fluctuations becomes particularly strong if the amount of molecules involved in a reaction is small. Here lies a difficulty of stochastic modeling. The boundary between a system where fluctuations in the species concentration are significant and a system in the macroscopic regime is not immediately apparent. It depends on relative timescales of processes in a biochemical network and the topology of the network, among other factors. Some of these issues we explore in Chapter 3 where we analyze a generic bacterial signaling network. The number of molecules sufficient to overcome diffusion limitation and to assure signal responses
1.1. Stochasticity fundamentals

Figure 1.1: Stochastic protein synthesis may leave individual cells with very few or none of the product if bursts in protein synthesis are infrequent and small within a cell’s generation. For the experimental evidence of this effect in *E.coli* see [Yu et al. 2006].

Accounting for stochastic effects in biochemical processes requires an entirely different take compared to a much more familiar macroscopic description where variables correspond to the amount (or the concentration) of chemical species. Since a fluctuating variable takes on many values with various probabilities, the theoretical description should not only allow for computation of the mean but also account for the second moment. A complete description should even track the evolution of the whole probability distribution, which prescribes how frequently values of the random variable (admissible by a particular random process) occur. As we shall demonstrate further, the shape of the steady-state probability distribution of the level of a biochemical product, a protein for instance, can reveal valuable information about the underlying stochastic mechanism. It may also imply physiologically meaningful behavior of cells.

In the recent decade in a number of settings, experimentalists have been able to measure distributions of protein concentration taken as snapshots over entire cell population [Elowitz et al. 2002] as well as temporal changes of a fluctuating protein. Theoretical arguments have been made regarding the equivalence (or lack thereof) between the two averages [Tănase-Nicola & ten Wolde 2008]. According to the ergodic principle, these two should be equivalent. However, the nature of the cellular system itself may prevent the evolution of all degrees of freedom from reaching all possible states in the lifetime of the organism. This may happen if fluctuations are slow (Fig. 1.1). Hence, not only the stochastic model requires one to study the whole spectrum of fluctuations instead of the mean, but also the knowledge of statistics of individual events is required to understand the physics of intracellular processes. Since it is difficult to infer about the physiology of the single cell based on statistical claims about population averages, we shall extend the analysis of cellular stochastic processes with the study of temporal properties of random events. We will scrutinize the steady-state distribution of waiting times between biochemical events. Owing to advancements in single-cell, single-molecule experimental techniques these type of theoretical studies are increasingly gaining solid experimental evidence.

There exist many equivalent approaches to model stochasticity at the mesoscopic level (e.g. master, Langevin, Fokker-Planck equations). For reasons explained earlier, one particularly didactic way of looking at fluctuations is by considering waiting times. A generic
Chapter 1. Introduction

Figure 1.2: Stochastic production-degradation model. (A) Schematic representation of the model. Chemical species $X$ is synthesized at the rate $k_p$ and degraded at the rate $k_d \cdot x$, where $x$ denotes the number of $X$ molecules. The inverse of parameters $k$ has an intuitive interpretation; $1/k_p$ is the mean interval between production events, $1/k_d$ is the mean lifetime of a single $X$ molecule. (B) Sample stochastic time trajectory. Synthesis and degradation events are drawn from the exponential waiting time distribution. (C) Stochastic birth and death of individual molecules $X$.

Birth-and-death model depicted in Fig. 1.2A includes only two stochastic processes. On average, the synthesis generates $k_p$ elements of $X$ per unit time. $X$ can represent a species of a bio-molecule like a protein or a messenger RNA, for instance. In the latter of the two reactions every molecule $X$ undergoes degradation. The average lifetime of such a molecule (the time between its synthesis and degradation) equals $\tau_d = 1/k_d$ time units.

If both, synthesis and degradation were taking place at exactly equal instances of time, the amount of molecules $X$, denoted by lower-case $x$, would remain constant at any given time (Fig. 1.3A). An instructive analogy would be a queue at the cash register. If a next customer joins (synthesis) the waiting people precisely at the moment the first person finishes paying (degradation), the length of the queue will not change. Unlike this deterministic situation, intervals between subsequent synthesis and degradation events are not fixed in a chemical system (nor in a real-life queue on a busy day in the supermarket). Precise timing of events is heavily influenced by thermal fluctuations affecting the molecules’ relative position which in turn influences their propensity to overcome the free energy barrier and undergo a chemical reaction.
Figure 1.3: The effect of waiting time statistics on steady-state product distribution in a linear birth and death process. (A-F) Every panel consists of three subplots: the top-left contains a normalized histogram of intervals (waiting time distribution) in the synthesis process with the mean of 5 events per unit time, the top-right contains an analogous histogram for the life-time of a single molecule $X$ – on average 1 event per unit time. The bottom graph depicts a normalized histogram of the amount of $X$ in steady-state (steady-state $X$ distribution). The solid curve in every panel depict the Poisson distribution – the distribution resulting from exponentially distributed synthesis and degradation processes (panel F).
Since the length of intervals between synthesis events and the lifetime of single molecules fluctuate, the steady-state number of $X$ will change in time accordingly. Figs 1.2B and C illustrate a sample time-course. But fluctuations of waiting times may differ in origin and hence may have a different statistic. Processes such as mRNA or protein degradation, assembly of protein initiation complex, enzymatic reactions, diffusion, they all involve a series of steps to complete. While the average behavior can be conveniently characterized by a model where the sequence of events is aggregated into a single step with an effective mean completion time, higher moments are sensitive to the model’s granularity. This is best illustrated in Fig. 1.3 where the steady-state $X$ distribution has the same mean regardless of the production and the synthesis statistics, however its standard deviation varies significantly. We shall see more examples of this in sections below.

How does the statistics of waiting times for synthesis and degradation events affect the magnitude and frequency of steady-state fluctuations of $X$? As explained earlier, synchronized additions and subtractions of the product leave the level of $X$ unchanged (Fig. 1.3A). The moment the synthesis intervals are no longer equally-timed, the degradation does not coincide with the production and the level of $X$ fluctuates in the course of time. The system becomes stochastic. Instead of a single value as in the deterministic case, $x$ may take a whole range with different probabilities. If the amount of $X$ synthesized balances the amount of $X$ degraded during the same period, the system is said to be in steady-state. Then, the distribution of $x$ can be plotted in a straightforward manner by creating a normalized histogram from equally-timed measurements of $x$. Clearly, along with the increased randomness of waiting times, the amount of states reached by $x$ rises and the steady-state product distribution widens (Fig. 1.3A-C). One waiting-time distribution is particularly privileged, however. Intervals between synthesis events drawn from the exponential function ($k_p \exp (-k_p t)$) result in the steady-state product distribution independent of the type of the waiting-time distribution for the degradation (Fig. 1.3D-F)! This phenomenon is known in queueing theory as insensitivity. Moreover, for all of these cases the steady-state $X$ distribution is Poissonian – the distribution being a seminal result for the birth-and-death model with exponential waiting times.

The special status of the exponential waiting-time distribution becomes clear if one attempts to characterize a stochastic process mathematically. A consistent description resulting in the evolution equation – the master equation (a stochastic counterpart of the ordinary differential equation), requires a so called Markovian assumption. The assumption implies that the future evolution of a stochastic process depends solely on the current state of the system and that no memory of previous events exists – the stochastic process is memoryless. Take the synthesis of $X$, for example. Suppose that the last production event took place one second ago. According to the memoryless assumption, the probability that the next event occurs after an additional two seconds, given that one second of waiting has already passed, equals just the probability of an event taking place after two seconds regardless of the amount of waiting before. Using a bus stop analogy: the probability of a bus arriving after two minutes would be the same for newcomers as for passengers already waiting at the bus stop. On average, one would wait for a bus for the same amount of time, and it would not matter when during the period between two bus departures one had arrived. It contradicts common sense; people arriving later after the last bus departure feel that the bus should appear sooner. This is because real buses run
(usually) according to a schedule and intervals between their departures obey a different statistics!

The memoryless assumption constrains the type of the function from which the intervals between events are drawn. It turns out that the only function satisfying this condition is the exponential. This has far-reaching consequences. Any realization of a memoryless stochastic process (a so-called Markov chain) implies waiting times between events being drawn from the exponential distribution (just like synthesis and degradation in Fig. 1.3F).

The time evolution of such a process has a tractable analytical representation in the form of the master equation. In fact, stochastic processes as those depicted in Figs 1.3A-C where uniform waiting time distribution are used have no closed analytical solutions. Approximations are required which complicates the analysis and numerical simulations.

In the limit of large numbers of reactants the memorylessness is also reflected in the Law of Mass Action. Consider a volume $V$ in which a linear decay of $N_0$ molecules of species $X$ takes place ($X \rightarrow \emptyset$). The dependency of the continuous concentration $[x] = x/V$ on time is easily obtained as $[N_0] \exp(-k_d t)$ from which the inverse equation, the time to reach a given concentration $[x]$, follows:

$$t_x = \frac{1}{k_d} \cdot \log \frac{[N_0]}{[x]}.$$  \hfill (1.1)

In the stochastic description, the time to decay for a single molecule is distributed exponentially with an average $1/k_d$, i.e. $\exp(-k_d t)$. The probability for the first decay event in an ensemble of $N$ molecules, can be computed by multiplying probabilities:

$$Pr(T^{(1,N)} > t) = Pr(1\text{-st molecule out of N decays after } t)$$
$$\equiv Pr(\text{all N molecules decay after time } t)$$
$$= Pr(T^1 > t \text{ and } T^2 > t \text{ and } \ldots \text{ and } T^N > t)$$
$$= Pr(T^1 > t) \cdot Pr(T^2 > t) \cdot \ldots \cdot Pr(T^N > t)$$
$$= e^{-N k_d t}.$$  \hfill (1.2)

The resulting probability for the first degradation is also exponentially distributed, but with a smaller average, $1/(Nk_d)$. By calculating the probability for the second, third and $k$-th degradation one obtains the mean time after which $x = N - k$ molecules are left in the system, i.e. the stochastic equivalent of Eq. 1.1:

$$\langle t \rangle_x = \frac{1}{k_d} \sum_{i=(N-k)+1}^{N} \frac{1}{i}.$$  \hfill (1.3)

For large $N$ these two equations (Eq. 1.1 and Eq. 1.3) are equivalent (Fig. 1.4).

The fact that the product of exponential functions is also an exponential greatly benefits the computational procedure. In a decay example, one (lengthy) algorithm would track the lifetime of every molecule. At the initial time zero, decay times would be drawn from the exponential distribution with the mean $k_d$ for each of $N$ particles. After ordering these next-decay times, the sequence of single-molecule subtractions from the ensemble would yield a sample time-dependent realization of the stochastic process. Such a trajectory could be obtained in a much simpler way with help of Eq. 1.2. The time for the first degradation event in a pool of $N$ particles can be drawn from the exponential distribution
Figure 1.4: Comparison of the stochastic and the deterministic model of a decay reaction, $X \rightarrow \emptyset$. Initially, the system contains 10 (left panel) and 100 (right panel) molecules of species $X$, each having the same average lifetime $\tau_X$. Stochastic mean (empty circles) is taken over 1000 trajectories simulated with the Gillespie algorithm. The discrepancy between the continuous model (dashed line) and the stochastic mean is apparent only for a small number of molecules (left panel).

with the decay constant $N \cdot k_d$. The exponential waiting time distribution guarantees the memoryless property of the process, thus the time for the next degradation is given by the same distribution but parameterized by $(N - 1) \cdot k_d$, and so forth (Fig. 1.5). Contrary to the former scheme, the resulting algorithm does not scale with the number of molecules, $N$; in order to determine the time of the next decay reaction, only one random number generation is required. Both of these approaches to numerically simulate chemical reactions have been considered in the past and they form the basis of a next-reaction and a direct method [Gillespie 2007, Gibson & Bruck 2000].

The above holds only for exponentials. Any other form of the waiting-time distribution complicates significantly the scaling of the average first event time (the mean first-passage}

Figure 1.5: A linear decay model with exponential waiting times. One method to create a stochastic trajectory is to draw the next decay time from the exponential waiting time distribution, $k_d \exp(-k_d t)$. When ordered, these times form a sequence of decay events. Such an algorithm scales with the number of molecules. Alternatively, one can take the advantage of Eq. 1.2. First, the time of the first decay is drawn from the exponential distribution with the mean $N \cdot k_d$, then the next decay time is parameterized by $(N - 1) \cdot k_d$, and so forth.
1.2 First example

As a result, the numerical simulation of such a system requires a much more involved implementation and (usually) more computational time. We shall come back to this point later.

Having in mind the severe constraint on the waiting time function imposed by the memoryless assumption one can appreciate the insensitivity of the steady-state product distribution to the type of the waiting time function describing degradation (Figs 1.3D-F). This result extends significantly the amount of solvable models and has a few interesting biologically relevant ramifications. Take mRNA or protein degradation, for instance. Both of these processes proceed in a sequential manner: the poly-adenylated tail of messenger RNA is removed step by step before the final degradation, attachment of ubiquitin monomers to the protein seals its fate and makes it recognizable by proteasomes [Pedraza & Paulsson 2008]. The waiting time statistics of sequential processes departs significantly from the exponential (a peaked, Gamma-like distribution). The insensitivity principle guarantees, however, that the steady-state distribution remains the same (Poissonian) as long as the waiting time distribution of the synthesis process is exponential.

A lot of intuition about biological systems relies on the memoryless property. In the following section we shall pursue some questions of central importance for understanding the origin of exponential approximation, the relationship between the waiting time distribution and steady-state properties such as the mean and noise. Turning to simple models should ease the exposition.

1.2 First example

The model from Fig. 1.2 is the simplest biochemical network with stochastic synthesis and degradation. In numerous biologically important cases, however, synthesis is affected by additional processes which, if collected into a single constant production rate $k_p$, would reproduce correctly only the average. One of such examples is activation of a gene by a transcription factor. It involves frequent association and dissociation events (a biomolecule makes contact with a site on DNA, forms a complex whose stability depends on the affinity of the binding site, and unbinds) interrupted by long periods during which the transcription factor engages in a long diffusive departure from the cognate DNA site. As a consequence, synthesis takes place only during those periods when the activating protein is bound; protein production occurs in bursts.

A coarse-grained model of such a switch-like product synthesis depicted in Fig. 1.6 omits details of association and dissociation events. At the microscopic level, these are typically affected by diffusion, electrostatics, hydrodynamic interactions or conformational changes of macromolecules. Nonetheless, this simplified description provides some valuable insights into how bursts of synthesis, and fluctuations in waiting times in general, affect the steady-state statistics. This relationship is largely unexplored for biochemical networks as most studies on the consequences of stochasticity for biological systems have focused on the noise in the steady-state level of molecular intermediates.

We assume that transitions between active and inactive states follow the exponential statistics and that production events in the active state occur at exponentially-distributed intervals. This might not hold true in general. A more detailed description could take into account a multi-step nature of gene activation: in order to render a gene active, a few proteins have to assemble a complex. This is the case in higher organisms.
Figure 1.6: The simplest model of bursty production. (A) Synthesis of species \( X \) is modulated by a switch. (B) Accumulation of \( X \) in time (degradation is omitted here). Light regions correspond to the OFF state, grey represent the active ON state. (C) Synthesis of \( X \) takes place during the ON state only.

[Degenhardt et al. 2009] and hence switching from OFF to ON state has to be described by a non-exponential (Erlang-type) peaked distribution (the topic explored in more detail in Chapter 2). For simplicity we shall focus on a simpler case where waiting times of all four transitions in Fig. 1.6 are described by the exponential distribution.

How is the overall waiting time distribution for the production events affected by the switching between synthetic activity and silence periods? In other words, what kind of probability distribution describes the statistics of product appearance given the silence periods interrupting an otherwise constant synthetic streak? Note that we assumed the lifetimes of the ON and OFF states and the intervals between production events in the ON state to have exponential distribution.

In case of a large time scale separation between the ON-OFF switch and the synthesis, i.e. many products are synthesized in the active state and the active state is comparably long to the inactive period, mostly two types of intervals are abundant: short ones corresponding to production events in the ON state, and long ones due to OFF periods (Fig. 1.6). Hence, the resulting distribution must consist of two differently-parameterized exponentials with weights depending on the time contribution from both states:

\[
    f(t) = w r_1 e^{-r_1 t} + (1 - w) r_2 e^{-r_2 t}. \tag{1.4}
\]

This heuristically derived equation is in fact an exact solution of the network in Fig. 1.6 for all time scales. The (non-trivial) relation of \( r_1, r_2 \) and \( w \) to kinetic parameters of the burst model in Fig. 1.6A can be found in the Appendix of Chapter 2.

We established that the stochastic production modulated by the ON-OFF results in the double-exponential waiting-time distribution. The simple birth-and-death model considered in Fig. 1.2 where synthesis events were spaced at exponentially distributed intervals produced a Poisson distribution of the product in the steady-state. Finally, we are able to address the question posed at the beginning of this chapter: how does this steady-state distribution change once the synthesis takes a different form, in this case the form of Eq. 1.4?
1.2. First example

Sample analytical solutions for steady-state product distribution are depicted in the frequency plot in Fig. 1.7. All of the distributions have the same mean, whereas their shape differs depending on the size of synthesis bursts and their significance. Notably, for small burst size the distribution may take the bell-curve-like shape centered around the mean or, when bursts increase in size, it may peak twice: around the base state (zero) and the steady-state mean in the active state given by the ratio $k_p/k_d$. The existence of such a bimodal distribution is in no way reflected in the macroscopic description; the stochastic model reveals a whole new family of behaviors.

1.2.1 Physiological implications of bursts

Living organisms employ numerous strategies to strive in their environment. Spontaneous genetic mutations accompanied by the selection process lead to fixation of new, more “successful” genotypes. However, the time scale of these changes is too long to adjust to changes in nutrient abundance taking place on an hourly basis, for instance. A multitude of sensory networks capable of inducing and regulating gene expression have evolved to cope with changing conditions. Even for bacterial sensing largely relying on relatively simple two-component networks, a high number of such networks (typically around few dozen per cell) and the fact that they operate in parallel with numerous cross-talks have led some to postulate a rudimentary form of intelligence embedded in the sensory network [Hellingwerf 2005]. Whether their complexity is sufficient to exhibit neural network-like characteristics such as memory or learning is an intriguing concept awaiting further theoretical and experimental evidence.

While the ability to “intelligently” process environmental inputs by bacterial sensory machinery remains a hypothesis, the ability to reflect temporal pattern of extracellular conditions in the structure of regulatory network has been recently demonstrated experimentally [Mitchell et al. 2009]. Two model organisms, *E.coli* and *S.cerevisiae*, used in this study evolved to be capable of activating parts of its machinery in anticipation of the sequence of stimuli. Random, unpredictable fluctuations in environmental conditions pose a grander challenge for organisms, however. Likewise, rapid but infrequent (on
Figure 1.8: Bimodal behavior of the steady-state product distribution may arise in a nonlinear dynamical system as well as in a stochastic switch. (A) Spontaneous stochastic switch, a “burster”, with constitutive production of P. If ON and OFF states are long enough to establish their respective quasi steady-states, the system will “flip” between two distinct states. (B) A classical mechanism leading to bimodality – a positive feedback. Product M promotes its own synthesis by forming a dimer D which further induces the active ON state. Dimerization introduces nonlinearity which may lead to bistability if none of the processes acts too strongly [Isaacs et al. 2003].

Cell’s generation scale) cues inducing irreversible lethal changes require a different type of response system. One solution is to maintain a multitude of regulatory mechanisms continuously prepared to face a vast array of environmental challenges. But even then, a cell fully equipped with sensory networks activating appropriate genes may respond to a change only if integration of the extracellular signal lasts long enough to average out noise inherent in the sensing procedure. If the time required for gathering sufficient information about the environment exceeds the generation time of a single cell, such a cell is never capable of reacting properly, which compromises the fitness of the population as a whole. Diversification of phenotypes may be a useful strategy to overcome this limitation.

Fitness advantage of a phenotypically heterogeneous population exposed to fluctuating conditions depends on the relative time scale of intracellular processes and changes in environmental cues [Kussell & Leibler 2005]. The choice of the strategy arguably depends on cells’ ability to gather information about the outside world. Cells switching their phenot-
Figure 1.9: Population heterogeneity increases survival rate. In the simplest case, the abundance of the protein dealing with environmental challenge (e.g. neutralizing a toxin) has a bimodal distribution due to stochastic gene expression or due to feedback regulation, a so-called feedback-based multistability [Smits et al. 2006]. The example of the latter mechanism could be a positive feedback in the sporulation network of \textit{Bacillus subtilis}. More generally, bi- or multi-modality could refer to meta-states where gene expression profiles of many genes undergo significant changes simultaneously [Chang et al. 2008]. Bimodality results in a stable, phenotypically heterogeneous population. Only individuals “prepared” for the upcoming change in the environment survive. The initial heterogeneity is recovered after few generations.

Types with the same frequency as changes in the environment have a greater survival rate as opposed to populations staying out of sync with the surroundings [Acar et al. 2008]. This observation hints at a possible choice of strategies aiming for adaptation. In environments with rapid, irregular or extreme changes the cost of maintenance of elaborate sensory networks may exceed the benefits of a higher survival rate. Additionally, the response of the network to a rapid change may be unsatisfactorily slow thus promoting a simpler solution. An isogenic (having the same set of genes) population could, for instance, increase its fitness by generating subpopulations by stochastic modulation of gene expression; each of these subpopulations performs sub-optimally in the “average” environment but is able to survive critical changes – a strategy known as bet-hedging (for a comprehensive review see [Davidson & Surette 2008]).

Bimodality in the gene product level as depicted in Fig. 1.8 is one of the elementary mechanisms behind induction of population diversity in a population of isogenic cells. Two peaks of the steady-state product distribution have been traditionally attributed to bistability in the dynamical system resulting from some form of feedback regulation; a so-called feedback-based multistability. A thorough review concerning this type of mechanisms can be found in [Smits et al. 2006]. Here, we discuss a different case where two
distinct levels of a protein across the population emerge solely due to stochastic gene regulation (Fig. 1.8) [Kepler & Elston 2001, To & Maheshri 2010]. Cellular heterogeneity due to bursty protein production may arise only if the frequency of ON-OFF transitions is comparable to (or slower than) the generation time of the cell (Fig. 1.1); fluctuations are “slow” [Sigal et al. 2006]. Only then, extensive inactivity periods during protein synthesis may leave some cells in the population lacking these molecules, while other cells may contain the entire production burst [Yu et al. 2006].

The emergence of stochastic bet-hedging strategy in a bacterial population has been recently observed in an elegant experiment [Beaumont et al. 2009]. After 15 rounds of subjecting cells to two opposing environments, each favoring a different phenotype, a genotype evolved capable of stochastic switching between the two conditions. Stochastic gene expression has been also shown to confer fitness advantage in a population of yeast cells exposed to antibiotic stress [Blake et al. 2006]. Due to bursts in the production of the protein conferring resistance to an antibiotic, (at least) part of the population is in the position to respond rapidly (Fig. 1.9). This translates to an overall increase in the survival rate of the whole population. Survival upon antibiotic treatment in general has been attributed to population heterogeneity stemming from stochastic switching between two phenotypes with distinct survival rates, a phenomenon known as bacterial persistence [Balaban et al. 2004, Kussell et al. 2005, Bishop et al. 2007]. Similarly, Sorger and colleagues [Spencer et al. 2009] have demonstrated how cancer cells escape drug treatment thanks to stochastically induced cell-to-cell variability. Studies like this should help answering a long-standing question of why seemingly identical cells respond differently to a drug? In a remarkable experiment of Cohen et al., levels and locations of approximately 1000 endogenous proteins have been tracked in individual cells after admitting a chemotherapy drug [Cohen et al. 2008]. The presence of the drug evoked a higher variability in protein levels and in some cases concentrations exhibited a bimodal distribution. The low and high levels in protein concentration corresponded to survival or death of a cell. Thus, the presence of the environmental stress induced a stochastic strategy, which allowed part of the cancer cell population to escape their deadly fate. Phenotypic diversity in a clonal population is rarely a result of fluctuations in the expression of a single gene, however. Cellular states with distinct functionalities (phenotypes) usually correspond to transcription profiles differing in the expression of many genes. Slowly-decaying fluctuations may promote reversible transitions between such meta-states implying various cell fates as has been demonstrated for mammalian progenitor cells [Chang et al. 2008].

Diversification of the microbial population benefits its survival rate upon environmental changes but also allows these simple organisms to solve complex tasks. In one of such systems, a population of Salmonella bacteria split stochastically into two subpopulations. One of the phenotypes facilitates infection in the gut lumen by triggering the inflammation. By doing so it contributes for the “public good” of the whole bacterial invasion, however, these bacteria get eliminated by the host’s immune response [Ackermann et al. 2008]. This self-destructive cooperation elegantly illustrates the population benefit of phenotypic noise and its role in bacterial pathogenesis. It is also an interesting contribution to recent experimental studies on the evolution of cooperation by engineering simple bacterial ecologies [Santorelli et al. 2008, Gore et al. 2009, Khare et al. 2009].
1.3 Timing of biochemical reactions

Thermal fluctuations, diffusion in crowded cellular environment, small numbers of the reactants, all of these affect rates of chemical reactions and contribute to the overall stochasticity in biochemical processes. Such inherent randomness may be selected in an evolving system if it increases the fitness of the organism [McAdams & Arkin 1999]. But living organisms are not falling apart, nor is their functioning completely erratic. Quite the contrary, their survival requires transfer of genetic information between generations, protection of this information against damages, processing of intracellular and environmental queues. Cellular events such as cell cycle, prediction of day-night rhythms, or anticipation of sequential environmental changes [Mitchell et al. 2009] require temporal synchronization between various processes.


1.3.1 Sequential processes

For a single molecule noise, or variance scaled with the squared mean, in the time to complete a biochemical transition (the waiting time) decreases if the transition is divided into substeps [Li & Qian 2002, Pedraza & Paulsson 2008]. If the timing of all \( N \) processes in the sequence is of similar duration, the resulting waiting time distribution becomes narrower than the waiting time distribution of a single-step transition (Fig. 1.10). For a simplified case where the rates of all intermediate reactions are equal, the noise depends inversely on the length of the sequence.

A process with events drawn from a non-exponential waiting time distribution does not have a straightforward counterpart in a macroscopic model described by ordinary differential equations (ODEs). Macroscopic biochemical models assume exponential waiting times, i.e. a Markov chain is the underlying stochastic description of chemical reactions (see Eqs. 1.1 and 1.3). Divergence from the memoryless distribution has a major consequence for the steady-state product distribution (Fig. 1.3) but also for the transient as illustrated by the linear decay reaction in Fig. 1.11. If the lifetime of a molecule is distributed according to a non-exponential function, in this case a Gamma distribution arising due to a step-wise degradation, the decay proceeds much faster than in an equiva-
Figure 1.10: Waiting time distribution (the first-passage time probability density function – FPT pdf) and noise ($\eta^2$, variance over squared mean) in a sequential process. Black dot indicates the mean of the distribution. As an example we consider a decay reaction of species X. (Left) A single molecule X decays in a single transition with rate constant $k_d$. The corresponding waiting time distribution is an exponential with mean and noise equal 1. (Middle) A single molecule X decays in two substeps with rates $k_{d1}$ and $k_{d2}$, respectively. Each substep has an exponential waiting time distribution. The overall distribution is a result of a convolution of the two distributions of the substeps. In this case, both rates equal 1, which results in the minimal noise and the narrowest distribution for such a sequence. (Right) Time scales of the two processes differ ($k_{d1} = 1$ and $k_{d2} = 0.1$). Noise approaches the value of the dominating step, i.e. the one with the larger mean.

The above discrepancy between the time evolution of one-step and multi-step decay processes illustrates an important modeling issue. Very frequently, a description of biochemical reactions involves “bundled” steps which in fact consist of a number of shorter steps. Formation of a transcription initiation complex, mRNA or protein degradation, signal transduction are just a few of such processes. As we have seen, a more detailed description accounting for these sub-steps has a significant effect on the steady-state noise or the time evolution of the system. Notably, in the latter case a mere expansion of the macroscopic model with additional linear reactions describing all the steps in the process will not match the stochastic transient. Reaction rates in such an expanded model would require intricate dependencies on state variables.

Timing of biochemical events becomes more precise if processes are aligned in a sequence. However, the time to complete such a sequence increases with the amount of steps. Therefore, a gain in noise reduction may be offset by an extended duration of the process and possibly a higher cost of production and maintenance of proteins that make up the reaction chain. The simple exit time process from Fig. 1.12 provides an illustration of these tradeoffs. Molecules of species X are synthesized and degraded in first-order reactions. Further action is triggered only if a certain amount of X is accumulated; thus,
1.3. Timing of biochemical reactions

Figure 1.11: The effect of the peaked waiting time distribution on a decay reaction. (A) Waiting time distributions for a Poisson process (exponential distribution - solid line), the sum of two exponentially distributed random variables (gamma distribution with \( k = 2 \) steps - dotted), and the sum of five variables (gamma with \( k = 5 \) steps - dashed). All distributions have the same mean equal to 1. (B) Stochastic dynamics of a decay reaction, \( X \rightarrow \emptyset \). Waiting times for a decay of every \( X \) molecule are drawn from the exponential distribution (solid line) and from a 5-step (\( k = 5 \)) gamma distribution of the same mean equal to 1 (dashed). Averages are taken over 100’000 trajectories. Additionally, three arbitrary stochastic trajectories are drawn for the exponential case (light solid).

The higher the threshold in the exit time process, the longer the time to reach it; the effect enhanced by the degradation which forces the build-up level down to the initial zero value. The width of the exit time distribution becomes narrower with the increasing threshold level only to widen again in the regime where the threshold is comparable to or exceeds the steady-state mean of the synthesis-and-degradation. The narrowest exit time distribution corresponds to the smallest variability coefficient, or noise; the timing for this threshold value is the least stochastic.

An example of a biological process described by this model is the switching of the rotation direction of the flagellar motor in bacteria such as \( E.coli \). Microbes are typically equipped with several independent motors. Their coordinated rotation is necessary to propel attached flagella and to direct the organisms towards regions more abundant with nutrients. How can a set of uncoupled motors can start their motion synchronously? One plausible explanation could be the threshold mechanism and its timing properties. A change in extracellular nutrient concentration gradient triggers a release of the activated signaling protein (phosphorylated form of CheY) at the cell’s “nose” – a relatively small region at one end of the cell (Fig. 4.3). Sequential build-up of these proteins at the motor site is required to change the direction of the flagella rotation. As shown in Fig. 1.12, the timing might be very precise if the activation threshold is placed slightly below the steady-state concentration. Consequently, a simultaneous switch of the coordinated motion can arise without any physical coupling of the motors (sic!)
Figure 1.12: Reaching the threshold in a simple synthesis and degradation reaction. (A) An illustration of a single stochastic trajectory crossing the threshold level of molecule X. (B) Distributions and noise in the time to reach the threshold value of X. The steady-state average in the synthesis-degradation model equals 20; $k_p = 20 \left[1/T\right], \ k_d = 1 \left[1/T\right]$. The smallest noise corresponds to the most peaked waiting time distribution.

1.3.2 Diffusion-limited reactions

Another wide class of processes giving rise to non-exponential waiting time distributions is a diffusive encounter of two reacting molecules (Fig. 1.13). In fact, diffusion can be perceived as a more general form of a sequential process where an infinite number of possible diffusive trajectories (sequences of small jumps in space) overlap. For a pair of initially separated molecules, the distribution of times to bind follows a peaked function similar to the distribution arising in the chain of chemical reactions.

The peakedness, and hence the smaller noise of the waiting time distribution can be further increased if many diffusing molecules perform the search simultaneously (Fig. 1.14). The increased precision in timing could be advantageous for signaling systems. When a change in extracellular conditions takes place, the reception of the signal by membrane sensors is proceeded by a race of a number of activated molecules (so called response regulators) towards the center of the cell. If one of the regulators successfully binds to a DNA site, gene expression can be altered in order to evoke appropriate physiological response. Due to a narrow waiting time distribution resulting from simultaneous diffusion of many molecules, the response will arguably vary only little among the cellular population allowing for a higher survival rate in face of a possibly fatal environmental change. However, mean and noise reduction is likely offset by the increased cost of protein production resulting, an issue investigated in detail in Chapter 3.

Dealing with non-exponential waiting time distributions for diffusion processes is cumbersome for only the Laplace form of the function is known analytically in most cases. Additionally, derivation of multi-particle distributions (according to Eq. A.13) requires knowledge of the explicit time-domain solution which often can be obtained only by means of numerical inversion. Many problems in biology, however, involve diffusive searches for small, compared to the entire volume of the cell, targets such as DNA binding site or a
1.4. Organization of the thesis

The generic burst-generating mechanism discussed in Fig. 1.6 was used to study statistics of intervals between syntheses of product $X$. In biological context, the ON/OFF switch represents the simplest model of gene activation and $X$ is the protein or messenger localized molecule. Such searches involve so many steps that the initial configuration has only a minor effect on the duration of the search. The process becomes effectively memoryless, and the waiting time distribution, although still peaked, is wide enough to allow for the approximation with the exponential function [Dobrzyński 2008] (see Fig. 1.14 – a single-particle waiting time distribution can be well approximated by an exponential function). As a result, variability of the time to reach the target (the first-passage time) increases and approaches that of the Poisson process (Fig. 1.15).

The memoryless approximation of search trajectories is a valuable simplification for modeling and simulations of spatially resolved problems. In this regime, a diffusion-limited reaction involving an encounter of two random walking biomolecules can be described by a first-order transition where the waiting time follows a memoryless exponential function. Costly computer simulations that are typically required to tackle problems involving diffusion may be therefore superseded by a much simpler model quickly solvable by kinetic Monte Carlo methods such as the Gillespie algorithm.

1.4 Organization of the thesis

The generic burst-generating mechanism discussed in Fig. 1.6 was used to study statistics of intervals between syntheses of product $X$. In biological context, the ON/OFF switch represents the simplest model of gene activation and $X$ is the protein or messenger localized molecule. Such searches involve so many steps that the initial configuration has only a minor effect on the duration of the search. The process becomes effectively memoryless, and the waiting time distribution, although still peaked, is wide enough to allow for the approximation with the exponential function [Dobrzyński 2008] (see Fig. 1.14 – a single-particle waiting time distribution can be well approximated by an exponential function). As a result, variability of the time to reach the target (the first-passage time) increases and approaches that of the Poisson process (Fig. 1.15).

The memoryless approximation of search trajectories is a valuable simplification for modeling and simulations of spatially resolved problems. In this regime, a diffusion-limited reaction involving an encounter of two random walking biomolecules can be described by a first-order transition where the waiting time follows a memoryless exponential function. Costly computer simulations that are typically required to tackle problems involving diffusion may be therefore superseded by a much simpler model quickly solvable by kinetic Monte Carlo methods such as the Gillespie algorithm.

1.4 Organization of the thesis

The generic burst-generating mechanism discussed in Fig. 1.6 was used to study statistics of intervals between syntheses of product $X$. In biological context, the ON/OFF switch represents the simplest model of gene activation and $X$ is the protein or messenger localized molecule. Such searches involve so many steps that the initial configuration has only a minor effect on the duration of the search. The process becomes effectively memoryless, and the waiting time distribution, although still peaked, is wide enough to allow for the approximation with the exponential function [Dobrzyński 2008] (see Fig. 1.14 – a single-particle waiting time distribution can be well approximated by an exponential function). As a result, variability of the time to reach the target (the first-passage time) increases and approaches that of the Poisson process (Fig. 1.15).

The memoryless approximation of search trajectories is a valuable simplification for modeling and simulations of spatially resolved problems. In this regime, a diffusion-limited reaction involving an encounter of two random walking biomolecules can be described by a first-order transition where the waiting time follows a memoryless exponential function. Costly computer simulations that are typically required to tackle problems involving diffusion may be therefore superseded by a much simpler model quickly solvable by kinetic Monte Carlo methods such as the Gillespie algorithm.
Figure 1.14: The effect of many random walkers (e.g. response regulators) on the first-passage time (FPT) pdf. Molecules start simultaneously at the outer membrane, $r_0 = R_{cell}$; the first-passage occurs when the first out of $N$ walker reaches the inner sphere of radius $a = 0.01R_{cell}$. For a single walker, $N = 1$, the FPT pdf corresponds to that in Fig. 3.8; the exponential approximation is indicated by black dots. Triangles denote the mean of the distributions. Mean and noise (not shown) in the waiting time decrease with the increasing amount of random walkers. (B) The mean first-passage time (MFPT) as function of the number of molecules for different target sizes $a$. Solid lines denote the exponential-like scaling $1/N$, which holds for small target sizes. Circles are computed using order statistics for the analytical result: numerical Laplace inversion of Eq. 3.5 to obtain a single particle FPT pdf, substitution of $f(t)$ into Eq. A.13, finally computation of moments of $f^{(1,N)}(t)$.

RNA (mRNA), the intermediate synthesized during protein production. In Chapter 2 we extend this picture and study how additional processes involved in protein synthesis influence the overall stochasticity of gene expression. We focus on a more realistic case, which includes transcription elongation, a process consisting of a multitude of small steps. We perform extensive computer simulations to demonstrate how transcription elongation can attenuate or enhance fluctuations introduced upstream of this process. In order to quantify our analysis we apply the first-passage time theory and introduce indexes for burst size, duration and significance, which have a clear analytical representation for the generic burst model. This chapter is based on published material [Dobrzyński & Bruggeman 2009].

Our focus on timing properties of stochastic events continues in Chapter 3. Here we look closer into a spatially-resolved system: a two-component signaling network. This relatively simple structure is responsible (among other tasks) for sensing changes in nutrient abundance outside of the cell and for inducing appropriate physiological responses within the cell. The response involves diffusion-limited searches of spatially-fixed targets by biomolecules. One of the main questions we set out to answer is how many molecules are required to facilitate a quick and robust response. Once again we employ the first-passage time theory and make use of the memoryless property of diffusion-limited reactions. Parts of the background theory used in this chapter has been published in form of conference.
Figure 1.15: Validity of Smoluchowski approximation in the model of target search. The mean and the noise are shown for the first-passage time of a molecule diffusing from the outer boundary at $R_{\text{cell}}$ (e.g. cell’s membrane) to a target (e.g. a DNA-binding site or cell’s nucleus) of radius $r_{\text{in}}$ in the center of a spherical cell. The target search depicted here is a model of the first step in the cellular signal response (discussed in more detail in Chapter 3). For targets much smaller than the cell’s radius ($r_{\text{in}}/R_{\text{cell}} \ll 0.1$) the mean search time is well approximated by the estimation derived from Smoluchowski diffusion-limited rate constant $K_D$, i.e. $\tau_D = 1/K_D = V_{\text{cell}}/4\pi D r_{\text{in}}$. Note that since noise is a dimensionless quantity it is also independent of the diffusion constant $D$.

Likewise, in Chapter 4 the analysis of the first-passage time distribution for a bimolecular reaction allows for a thorough comparison of spatial-stochastic computation methods. These methods are used in order to obtain statistics of biochemical processes involving diffusion. Higher moments and steady-state distributions of reactants are typically computed this way. We analyzed available methods which assume different underlying physical models of the diffusion and identified sources of discrepancies in their results. This chapter is based on published material [Dobrzyński et al. 2007].

Three appendices include mathematical definitions and derivations of essential equations discussed throughout the thesis. Appendix A covers basic concepts of the first-passage time theory such as the first-passage time probability density function, com-
putation of the first and the second moment of the distribution, and order statistics. Appendix B focuses on the first-passage time theory in the context of diffusion-limited reactions. Examples of derivations for one-dimensional geometries can be found there. The methods presented there along with the order statistics from Appendix A are used extensively in Chapter 3 to compute distributions and moments of first-passage times for many diffusing molecules in a signaling network. Derivations specific to this three-dimensional problem are included in Section 3.5 – Materials and Methods. Appendix C includes a detailed description of models and computational methods along with algorithms applicable to modeling of spatial stochastic problems in biology.